

UNIVERSITÀ DEGLI STUDI DI FIRENZE



Dipartimento di Statistica “G. Parenti”

Dottorato di ricerca in Statistica applicata

XXIII Ciclo - SECS-S/01

**Modelli ad equazioni strutturali multilivello per la  
valutazione dell’efficacia esterna del titolo di  
dottorato di ricerca**

Lucio Masserini

Tutor: Prof. Bruno Chiandotto

Coordinatore: Prof. Fabio Corradi



# Sommario

Sommario .....	i
Dedica .....	vii
Presentazione .....	ix
1. La valutazione dell'efficacia esterna .....	1
1.1 La valutazione dei corsi di dottorato .....	2
1.2 Lo strumento di misura proposto .....	5
1.3 Il modello proposto .....	9
2. Modelli ad equazioni strutturali multilivello con variabili latenti .....	13
2.1 Specificazione del modello e assunzioni .....	13
2.1.1 Il modello a variabili latenti .....	14
2.1.2 Il modello di misura .....	17
2.1.3 Il modello con variabili esogene osservate .....	20
2.1.4 La scomposizione degli effetti tra variabili .....	20
2.2 Le matrici implicate dei momenti .....	22
2.3 Modelli ad equazioni strutturali con variabili risposta categoriche .....	24
2.3.1 Modello di misura con variabili risposta latenti .....	25
2.3.2 Modello di misura con variabili esplicative osservate .....	30
2.3.3 Le matrici implicate dei momenti con variabili risposta latenti .....	31
2.4 La stima dei parametri .....	32
2.4.1 Stima con variabili osservate continue distribuite normalmente ...	33
2.4.2 Effetti con variabili osservate non distribuite normalmente .....	37
2.4.3 Stima con variabili osservate non distribuite normalmente .....	38
2.4.4 Effetti con variabili osservate categoriche .....	43
2.4.5 Stima con variabili osservate categoriche .....	44
2.4.6 Stima ad informazione limitata .....	46
2.5 La standardizzazione dei parametri .....	53

2.6	Adattamento del modello .....	54
2.6.1	La statistica test del modello .....	55
2.6.2	Analisi dei residui .....	57
2.6.3	Misure di adattamento comparativo o incrementale .....	58
2.6.4	Criteri di selezione del modello .....	62
2.6.5	Effetti con variabili osservate non distribuite normalmente o aventi modalità categoriche .....	65
2.6.6	Misure di adattamento con variabili osservate non distribuite normalmente .....	66
2.7	L'analisi multilivello .....	69
2.7.1	Aggregazione e disaggregazione delle variabili .....	70
2.7.2	Prospettive di analisi .....	72
2.8	Modelli ad equazioni strutturali multilivello .....	74
2.8.1	L'analisi disaggregata a due livelli .....	76
2.8.2	Il modello di misura a due livelli .....	78
2.8.3	La stima pseudo bilanciata .....	80
2.8.4	La stima di massima verosimiglianza ad informazione completa .	83
2.8.5	La stima con variabili osservate categoriche .....	85
2.8.6	Il metodo dei minimi quadrati ponderati ad informazione limitata	87
2.9	Modelli ad equazioni strutturali multilivello generalizzati .....	92
2.9.1	Il modello di risposta .....	93
2.9.2	La tipologia delle variabili risposta .....	94
2.9.3	Il modello a variabili latenti .....	96
2.9.4	La log-verosimiglianza marginale .....	97
3.	Il modello per la valutazione dell'efficacia esterna del titolo di dottorato di ricerca .....	101
3.1	La definizione delle variabili esplicative latenti .....	102
3.1.1	La misurazione della soddisfazione .....	103
3.1.2	La misura delle "Aspettative" e della "Partecipazione lavorativa".	104
3.2	La scelta del modello .....	106
3.3	Le caratteristiche del modello .....	108
3.4	Il sistema delle relazioni ipotizzate .....	110

3.4.1	Il sistema delle relazioni di primo livello .....	110
3.4.2	Il sistema delle relazioni di secondo livello .....	113
3.4.3	La rappresentazione grafica del modello proposto .....	116
3.5	La formalizzazione del modello .....	117
3.5.1	La misurazione delle variabili latenti .....	118
3.5.2	Il modello a variabili latenti .....	121
3.6	I risultati .....	122
3.6.1	L'analisi della struttura a due livelli .....	123
3.6.2	L'analisi del modello di misura .....	126
3.6.3	L'analisi delle relazioni di primo livello .....	128
3.6.4	L'analisi delle relazioni di secondo livello .....	129
3.6.5	Gli indici di adattamento del modello .....	130
3.6.6	La rappresentazione grafica del modello finale .....	132
3.7	Limiti del modello proposto .....	133
	Conclusioni .....	134
	Bibliografia .....	137
	Appendice A: La situazione occupazionale dei dottori di ricerca in Italia: una breve rassegna .....	145
	Introduzione .....	145
A.1	Panorama internazionale .....	145
A.2	Panorama nazionale .....	148
A.2.1	Iniziative MIUR e CNVSU .....	148
A.2.1.1	Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) e Ministero del Lavoro, della Salute e delle Politiche Sociali (MLSPS) .....	148
A.2.1.2	Comitato Nazionale di Valutazione del Sistema Universitario (CNVSU) .....	149
A.2.2	Iniziative di ateneo .....	151
A.2.2.1	Nuclei di Valutazione delle Università di Milano, Milano Bicocca e Trento .....	151
A.2.2.2	Nucleo di valutazione della LUISS .....	151
A.2.2.3	Nucleo di valutazione dell'Università di Bologna .....	151

A.2.2.4	Nucleo di valutazione dell'Università di Pavia .....	152
A.2.2.5	Nucleo di valutazione dell'Università di Camerino .....	153
A.2.2.6	Comitato Statistico dell'ateneo di Ferrara .....	154
A.2.2.7	Nucleo di Valutazione dell'Università Cattolica del Sacro Cuore di Milano .....	154
A.2.2.8	Servizio Formazione alla Ricerca e Servizio Ricerca dell'Università di Padova .....	155
A.2.2.9	Università di Siena .....	156
A.2.3	Iniziative di altri enti .....	157
A.2.3.1	Indagine sui dottorati di ricerca in Campania .....	157
A.2.3.2	Iniziativa Interuniversitaria STELLA (Statistiche sul TEma Laureati & Lavoro in Archivio on-line) .....	158
A.2.3.3	ISTAT .....	159
Appendice B. Condizione occupazionale dei Dottori di Ricerca: Questionario ad . un anno dal conseguimento del titolo .....		161
Appendice C. La condizione occupazionale dei dottori di ricerca del 2008: analisi descrittiva .....		181
Introduzione .....		181
C.1 L'indagine .....		181
C.1.1	La popolazione .....	182
C.1.2	La modalità di realizzazione dell'indagine .....	182
C.1.3	Il questionario .....	183
C.1.4	I dati raccolti .....	184
C.1.5	Il confronto di struttura .....	184
C.2 La condizione occupazionale dei dottori di ricerca del 2008 .....		186
C.2.1	La formazione ricevuta .....	186
C.2.1.1	Le motivazioni dell'iscrizione al corso di dottorato .....	186
C.2.1.2	La valutazione dell'esperienza formativa .....	187
C.2.1.3	L'esperienza lavorativa durante gli studi .....	188
C.2.1.4	Il periodo di studi .....	189
C.2.1.5	La condizione occupazionale attuale .....	189
C.3 Chi attualmente lavora .....		191

C.3.1	Il lavoro attuale .....	191
C.3.2	Tempi e modalità di ingresso nel mondo del lavoro .....	192
C.3.3	La tipologia del lavoro attuale .....	193
C.3.4	Le mansioni di ricerca ed innovazione .....	194
C.3.5	Il profilo giuridico del lavoro dipendente .....	195
C.3.6	Le caratteristiche dell'attività di ricerca .....	196
C.3.7	Le caratteristiche del contesto di lavoro .....	197
C.3.8	L'utilità sul lavoro della formazione ricevuta durante il corso di dottorato .....	197
C.3.9	Il livello di soddisfazione per il lavoro attuale .....	198
C.3.10	Il livello di retribuzione attuale .....	199
C.3.11	La ricerca di un nuovo lavoro .....	203
C.4	Chi attualmente non lavora ma ha lavorato dopo il conseguimento del titolo .....	202
C.4.1	L'ultimo lavoro e gli aspetti migliorati dopo il conseguimento del titolo .....	202
C.4.2	L'interruzione dell'ultima attività lavorativa .....	202
C.4.3	La ricerca di un nuovo lavoro .....	203
C.5	Chi non ha mai lavorato dopo il conseguimento del titolo .....	203
C.5.1	La ricerca del primo lavoro .....	203
C.5.2	I motivi della non ricerca .....	204





*A mio padre*



## ***Presentazione***

Il dottorato di ricerca è stato introdotto nel sistema universitario italiano, quale terzo livello di formazione superiore, nel 1980 dal DPR dell'11 luglio n. 382; la gestione dei dottorati era interamente demandata a decreti ministeriali.

Con l'introduzione dell'autonomia universitaria, si perviene ad un'abrogazione quasi integrale della normativa preesistente (legge del 3 luglio 1998 n. 210 e decreto ministeriale del 30 aprile 1999 n. 224), assegnando alle singole università la potestà di decidere sull'istituzione dei corsi di dottorato, sulla fissazione degli obiettivi formativi, le modalità di accesso, la programmazione, la ripartizione del numero di dottorati e di borse, la costituzione delle commissioni per la valutazione dei dottorandi e il conferimento del titolo.

A trenta anni dall'istituzione dei corsi di dottorato, e nella fase di transizione e trasformazione particolarmente delicata che sta attraversando il sistema universitario italiano, qualunque tentativo di valutazione dell'esperienza vissuta può fornire un utile contributo al processo di revisione dell'ordinamento in atto.

Questo lavoro di tesi, realizzato nell'ambito del progetto di ricerca finanziato dal Comitato Nazionale per la Valutazione del Sistema Universitario (*CNVSU*) ed affidato ad un gruppo di lavoro facente capo al Dipartimento di Statistica "G. Parenti" dell'Università degli Studi di Firenze, si colloca in tale ottica.

La tesi è articolata in tre capitoli. Nel primo capitolo si procede ad una disamina molto sommaria della problematica riguardante la valutazione dei processi formativi offerti dal sistema universitario italiano e alla definizione della misura di efficacia esterna che si propone di analizzare. Il secondo capitolo è dedicato alla trattazione dei modelli ad equazioni strutturali multilivello basata su una rassegna critica dei contributi teorico-metodologici più recenti presenti in letteratura. Nel terzo capitolo, dopo aver definito le variabili esplicative dell'efficacia esterna e formalizzato il modello, vengono illustrati i risultati conseguiti. Le conclusioni e tre appendici completano il lavoro di tesi. L'appendice A contiene una rassegna non esaustiva delle indagini riguardanti la condizione occupazionale dei dottori di ricerca, l'appendice B riporta il questionario mentre l'appendice C prevede un'analisi descrittiva dei dati.



# Capitolo 1

## La valutazione dell'efficacia esterna

### *Introduzione*

Il tema della valutazione interna nelle università in Italia è stato introdotto dalle leggi 168/89 e 537/93; la prima legge prevede l'attuazione di forme di controllo interno sull'efficienza e sui risultati della gestione nelle università; la seconda legge prevede l'istituzione nelle università dei Nuclei di Valutazione Interna (NVI).

Una nuova legge (L. 19 ottobre, n. 370 – G.U. n. 252 del 26/10/1999) dispone norme volte a disciplinare, più compiutamente, la valutazione del sistema universitario: *“Le università adottano un sistema di valutazione interna della gestione amministrativa, delle attività didattiche e di ricerca, degli interventi di sostegno al diritto allo studio ...”* (art. 1, comma 1). Ulteriori precisazioni sulla riforma sono contenute nel D.M. n.270/04 contenente le modifiche al regolamento recante le norme concernenti l'autonomia didattica degli atenei.

Una specificazione del *‘cosa e come valutare’* i Corsi di Dottorato è contenuta nel Regolamento in materia di Dottorato di Ricerca (D.M. 30 aprile 1999, n. 224, art. 3, comma 2) e meglio specificato nell'RDR 01/06: *“Progetto per la ricognizione, raccolta e analisi dei dati esistenti sul dottorato di ricerca e per l'indagine sull'inserimento professionale dei dottori di ricerca”* del Comitato Nazionale per la Valutazione del Sistema Universitario (CNVSU) e nella nota Ministeriale: *“Relazione 2009 dottorati di ricerca”* del 23 gennaio 2009.

L'obiettivo principale di questa tesi è quello di evidenziare le potenzialità ed i limiti della modellistica strutturale multilivello, sia attraverso un'approfondita disamina critica dei contributi teorico-metodologici presenti nella letteratura più recente, sia attraverso l'introduzione di uno specifico modello ad equazioni strutturali multilivello sufficientemente rappresentativo del fenomeno indagato e, quindi, capace di fornire spunti di riflessione che si possano rivelare di qualche utilità nell'attività di programmazione e gestione della formazione di terzo livello (dottorato) in Italia. Il modello ad equazioni strutturali multilivello sarà utilizzato per la valutazione dell'efficacia esterna del titolo di dottore di ricerca analizzando la condizione occupazionale dei dottori che hanno conseguito il titolo nel 2008, osservati ad un anno dalla conclusione del percorso formativo.

## **1.1 La valutazione dei corsi di dottorato**

Il dottorato di ricerca costituisce il terzo livello di formazione universitaria ed è il grado più alto di specializzazione offerto dalle università sia per le carriere accademiche e di ricerca sia per quelle del mondo produttivo, in particolare di quello attento all'innovazione. Negli ultimi anni il numero di dottori di ricerca in Italia è cresciuto significativamente, in conseguenza anche all'aumento dell'offerta formativa e le finalità del dottorato di ricerca si sono ampliate rispetto a quelle tradizionali.

L'analisi del contributo che il titolo di dottorato di ricerca ha nei confronti dell'inserimento occupazionale in ambito accademico o, più in generale, in quello del mondo del lavoro, rappresenta quindi uno strumento conoscitivo rilevante per la valutazione sia della qualità che dell'efficacia dei corsi di dottorato. Infatti, se al momento della sua istituzione il dottorato di ricerca aveva una valenza occupazionale circoscritta e finalizzata in modo quasi esclusivo all'attività di ricerca scientifica<sup>1</sup> da svolgere nella generalità dei casi in ambito accademico, le norme successive ne hanno ridefinito gli obiettivi introducendo un collegamento prima inesistente con il mondo produttivo<sup>2</sup>, fino a prevedere la possibilità che i dottorandi svolgano esperienze in

---

<sup>1</sup> Il dottorato di ricerca è stato previsto dal D.P.R. 382/1980 "quale titolo accademico valutabile unicamente nella ricerca scientifica" (art. 68).

<sup>2</sup> L'art. 4 della L. 210/98 che ha abrogato le normative precedenti stabilisce che "i corsi per il conseguimento del dottorato di ricerca forniscono le competenze necessarie per esercitare, presso università, enti pubblici o soggetti privati, attività di ricerca di alta qualificazione".

contesti lavorativi<sup>3</sup>.

La necessità di valutare la condizione occupazionale dei dottori di ricerca trova conferma anche in un recente documento della Conferenza dei Rettori delle Università Italiane (CRUI): “Criteri per la revisione del dottorato di ricerca”, secondo il quale, tra gli indicatori di qualità per la valutazione delle Scuole di dottorato, la “condizione occupazionale dei dottori di ricerca entro cinque anni dal conseguimento del titolo” assume il ruolo di “misura del raggiungimento dell’obiettivo principale del dottorato di ricerca”.

La conoscenza degli esiti occupazionali dei dottori di ricerca e del percorso di carriera nel mercato del lavoro diventa quindi essenziale e può aiutare a ridurre la distanza fra il mondo universitario e della ricerca ed il mondo economico-produttivo.

Il problema della valutazione della qualità della formazione universitaria si può inserire nello schema proposto inizialmente da Lockheed e Hanushek (1994) e, successivamente, generalizzato da Chiandotto (2004, 2008). Secondo tale impostazione, la stima della performance complessiva di un sistema formativo di terzo livello può essere scomposta in tre fasi distinte (cfr. Schema sotto riportato): la prima relativa alla valutazione del modo in cui le risorse vengono impiegate per ottenere il risultato desiderato (analisi di efficienza); la seconda relativa alla valutazione qualitativa di tale risultato e al grado di raggiungimento degli obiettivi previsti (analisi di efficacia); la terza relativa alla percezione soggettiva degli utenti circa il servizio erogato. Tale valutazione può essere effettuata con una visione interna o esterna al sistema formativo, a seconda che si presti attenzione ai risultati e agli obiettivi conseguiti rispettivamente quando i soggetti coinvolti si trovano ancora dentro il sistema universitario o invece siano al di fuori, tipicamente nel mercato del lavoro.

Dallo schema si desumono quindi otto modi diversi di misura della qualità dei processi formativi di terzo livello. Limitandosi all’analisi dell’efficacia esterna, si possono quindi definire due metodi diversi di valutazione, il primo fa riferimento a misure di natura “oggettiva” (termini fisici), il secondo a misure di natura “soggettiva” (livelli di soddisfazione).

Tra le misure oggettive per la misura dell’efficacia esterna si possono considerare:

---

<sup>3</sup> Il DM 224/99 “Regolamento in materia di dottorato di ricerca” stabilisce, tra i requisiti di idoneità delle sedi, “la collaborazione con soggetti che consentano ai dottorandi lo svolgimento di esperienze in contesti lavorativi”.

il tempo intercorso tra il conseguimento del titolo di dottore e prima occupazione; la tipologia dell'occupazione, se stabile o precaria; il riconoscimento del titolo conseguito; la coerenza del lavoro svolto e la formazione ricevuta; l'utilizzo delle competenze acquisite nell'attività lavorativa; il livello di retribuzione; ecc.. Le misure soggettive di efficacia esterna possono caratterizzarsi in termini di soddisfazione espressa dai dottori di ricerca nei confronti sia dei caratteri "oggettivi" sopra elencati sia di altri aspetti riguardanti l'attività lavorativa svolta (prospettive di guadagno e di carriera, autonomia, coinvolgimento nei processi decisionali, ecc.); ma, su queste problematiche si avrà modo di soffermare l'attenzione nelle pagine successive essendo proprio questo l'oggetto del lavoro di tesi.

*Schema di valutazione dei corsi di dottorato*

	INTERNA AL SISTEMA	ESTERNA AL SISTEMA
TERMINI FISICI	<b>Efficacia interna</b> <i>effetto del corso di dottorato sulla capacità di apprendimento del dottorando</i>	<b>Efficacia esterna</b> <i>effetto del corso di dottorato sulla capacità lavorativa del dottore di ricerca</i>
SODDISFAZIONE	del dottorando rispetto allo <i>insegnamento</i>	del dottore di ricerca rispetto alla <i>condizione lavorativa</i>
TERMINI MONETARI	<b>Efficienza interna</b> <i>analisi costi/ricavi aziendali dell'investimento</i>	<b>Efficienza esterna</b> <i>ritorno economico dovuto al corso di dottorato frequentato</i>
SODDISFAZIONE	del dottore di ricerca rispetto alle <i>risorse impiegate</i>	del dottore di ricerca rispetto alla <i>condizione economica</i>

Nello schema proposto, la soddisfazione per il lavoro svolto rappresenta una *misura indiretta* ed *ex-post* della qualità del servizio formativo. Al riguardo ci si deve chiedere se una tale misura possa davvero essere interpretata come indicatore dell'entità dell'incremento di "*capitale umano*" che la formazione universitaria è riuscita a determinare. L'università fornisce una formazione di livello elevato e se questa consente a chi la ha acquisita una vita soddisfacente, si può certamente concludere che la stessa università abbia assolto la sua funzione primaria che è quella di fornire competenze che siano "spendibili" efficacemente nel mercato del lavoro; ovviamente, la soddisfazione per il lavoro svolto, che può essere collocato tra gli indicatori più significativi di utilizzo



proficuo della formazione ricevuta, non è l'unico ipotizzabile e fornisce esclusivamente informazioni nell'ambito di uno specifico contesto di analisi. In proposito occorre, inoltre, sottolineare che un'eventuale insoddisfazione per il lavoro svolto è indice di uno sfasamento tra ciò che l'università offre ai suoi studenti in termini di bagaglio formativo e ciò che il mondo del lavoro richiede o offre in merito ai vari aspetti lavorativi (possibilità di guadagno, possibilità di carriera, coerenza con gli studi fatti, uso delle competenze acquisite all'università, ecc.). Riguardo alle cause di un eventuale sfasamento sarà necessario ricorrere ad adeguati approfondimenti dell'analisi per comprendere se le deficienze riscontrate provengono dal mondo universitario, nel senso di una incapacità di adeguarsi alle richieste del mondo del lavoro, oppure se le stesse sono ascrivibili a quest'ultimo, nel senso di mancato o non adeguato sfruttamento dei profili professionali elevati quali quelli posseduti da un dottore di ricerca. Naturalmente, in un caso o nell'altro i correttivi applicabili dovranno essere diversi.

## **1.2 Lo strumento di misura proposto**

Come già sottolineato, la tesi si concentra sulla valutazione qualitativa dei risultati conseguiti all'esterno del sistema universitario ed ha per oggetto l'analisi della condizione occupazionale dei dottori di ricerca. In particolare, l'obiettivo è quello di valutare il contributo che il titolo di dottore ha nei confronti della realizzazione professionale dei dottori di ricerca e si configura come un'analisi dell'efficacia esterna della formazione universitaria di terzo livello. In particolare, l'analisi riguarda la valutazione in termini di soddisfazione della condizione lavorativa ad un anno dal conseguimento del titolo e si basa sulla percezione soggettiva dei dottori di ricerca.

L'efficacia esterna si può considerare come il contributo degli elementi tipici di un processo formativo nei confronti della probabilità di successo individuale nel mondo del lavoro (sia in termini di conoscenze che di capacità professionali), al netto dei fattori individuali e di quelli economico-ambientali.

Sull'efficacia esterna della formazione universitaria incidono in maniera rilevante la "qualità" dei servizi formativi offerti (tipologia dei corsi, organizzazione della didattica, capacità del corpo docente, ecc...) ma anche e soprattutto i soggetti che usufruiscono dei servizi formativi (dottorandi). Questi si differenziano tra loro per

caratteristiche (età, genere, ambiente familiare, ecc...) e rappresentano al tempo stesso materia prima, risultato finale ed utenti del servizio in quanto partecipano attivamente allo svolgimento del processo, condizionandone i risultati.

In generale, non esiste una vera e propria misura di efficacia esterna universalmente riconosciuta, poiché si tratta di un concetto astratto non misurabile direttamente. Per questo motivo si propone di misurarla, almeno in prima approssimazione, ricorrendo ad una serie di indicatori osservati che si possono ragionevolmente ritenere a questa correlati. L'efficacia esterna del titolo di dottorato si può quindi considerare come una variabile latente<sup>4</sup>, i cui valori non sono osservabili (Jöreskog e Sörbom, 1979) ma la cui esistenza determina associazione tra i valori osservati degli indicatori considerati<sup>5</sup>.

Tali indicatori sono stati individuati tra le risposte alla domanda "A30" (cfr. Appendice B) in cui si chiedeva ai dottori di esprimere il proprio livello di soddisfazione riguardo ad una serie di aspetti ritenuti rilevanti. Per ciascun aspetto, il giudizio è misurato con una variabile categorica suddivisa in dieci modalità ordinate (1 = "per niente" e 10 = "moltissimo"). Ognuna di queste variabili si può considerare come un indicatore di una variabile latente sottostante, i cui valori esprimono su una scala continua modalità osservabili unicamente mediante la corrispondente variabile risposta categorica. Gli indicatori di efficacia esterna considerati sono:

- la "coerenza con gli studi fatti";
- l'"utilizzo delle competenze acquisite";
- la "rispondenza ai propri interessi culturali".

---

<sup>4</sup> In generale, non esiste una definizione unitaria di variabile latente nonostante l'utilizzo sia ormai molto diffuso in diverse discipline. Piuttosto, esistono numerose definizioni i cui significati derivano dai modelli statistici in cui le variabili latenti sono impiegate. Le variabili latenti hanno anche assunto denominazioni differenti a seconda dell'ambito disciplinare di applicazione. Tra queste, ne ricordiamo alcune, "variabili non misurabili", "fattori", "variabili non osservabili", "variabili ipotetiche", "variabili sottostanti", "costrutti ipotetici", "effetti casuali", "classi latenti" e "fragilità". Le variabili latenti sono impiegate per rappresentare variabili osservate misurate con errore, fattori comuni nell'analisi fattoriale, eterogeneità non osservata, dati mancanti, risultati potenziali e risposte latenti sottostanti a variabili categoriche. Per una trattazione più completa dell'argomento si può consultare Bollen (2002).

<sup>5</sup> In base a questa assunzione, gli indicatori osservati sono condizionatamente indipendenti data l'esistenza della variabile latente, cioè la dipendenza (o correlazione) tra gli indicatori è dovuta esclusivamente alla loro associazione con la variabile latente. Questa proprietà di indipendenza condizionata o "locale" costituisce la base della definizione di indipendenza locale delle variabili latenti (Lord, 1953; Lazarsfeld, 1959).

In pratica, a parità di altre condizioni, il titolo di dottorato di ricerca si ritiene efficace rispetto alla condizione lavorativa attuale se l'attività svolta è coerente con gli studi svolti, se le competenze acquisite sono effettivamente utilizzate e se l'attività lavorativa è in linea con gli interessi culturali dei dottori. Più specificatamente, l'efficacia esterna si ritiene maggiore nei casi in cui i diversi indicatori osservati assumono valori più elevati.

L'efficacia del titolo di dottorato rispetto alla condizione lavorativa attuale, misurata in termini di soddisfazione rispetto agli aspetti considerati, è la variabile endogena che si intende analizzare e sarà definita "Efficacia esterna", indicata con (E).

La relazione ipotizzata tra la variabile latente "Efficacia esterna" e gli indicatori osservati (modello di misura) è definita con un modello di analisi fattoriale confermativa (Jöreskog, 1969) e si può rappresentare graficamente nel modo seguente (Figura 1.1). L'ellisse rappresenta la variabile latente, i rettangoli rappresentano le variabili osservate mentre le frecce rettilinee direzionali che partono dalla variabile latente rappresentano la relazione ipotizzata tra questa e le variabili osservate. La freccia orientata verso le variabili osservate rappresenta invece l'errore di misura.

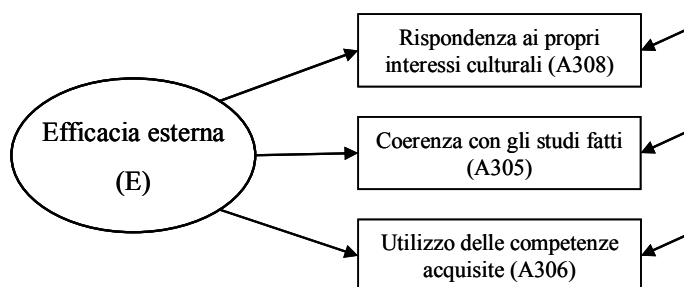


Figura 1.1. Relazione tra la variabile latente "Efficacia esterna" e gli indicatori osservati

La scelta di utilizzare questi indicatori definisce un'analisi dell'efficacia esterna del titolo di dottorato orientata a valutare gli aspetti "qualitativi" dell'attività lavorativa. Per questa ragione la misura dell'efficacia esterna non tiene conto di altri indicatori come ad esempio i tempi di inserimento nel mondo del lavoro ed il livello di retribuzione percepito poiché non forniscono alcuna informazione sulle caratteristiche e sulle qualità dell'attività svolta. Inoltre, considerato che l'analisi è riferita ai dottori di ricerca che hanno conseguito il titolo nel 2008, osservati dopo solo un anno, la variabilità dei tempi di inserimento nel mondo del lavoro è piuttosto contenuta. Per questo motivo, la sua

capacità esplicativa non sarebbe comunque in grado di fornire un contributo significativo alla spiegazione del fenomeno analizzato.

L'analisi è limitata ai dottori di ricerca che hanno conseguito il titolo nell'anno 2008 che attualmente lavorano (Blocco A del questionario - cfr. Appendice B). Sono quindi esclusi coloro che attualmente non lavorano ma hanno lavorato dopo il conseguimento del titolo (Blocco B del questionario - cfr. Appendice B) e coloro che non hanno mai lavorato dopo il conseguimento del titolo (Blocco C del questionario - cfr. Appendice B). La scelta di escludere dall'analisi coloro che attualmente non lavorano ma hanno lavorato dopo il conseguimento del titolo è motivata dalla volontà di limitare l'osservazione alle esperienze lavorative relativamente più durature. In questo modo si intende rendere più omogenea la situazione lavorativa iniziale ed evitare che le diverse condizioni di partenza possano avere un effetto significativo sulla valutazione dell'efficacia dei corsi.

I dati non sono di tipo sperimentale, poiché i dottori di ricerca non sono assegnati casualmente ai corsi di dottorato. La scelta del corso di dottorato frequentato può comportare il problema dell'autoselezione o *selection bias* delle unità (Maddala, 1983). Infatti, non si può escludere che esistano variabili non osservate (motivazioni, abilità, attitudini e capacità dei dottori) che abbiano influenzato la scelta del corso frequentato e di conseguenza, i valori delle variabili risposta osservate.

L'efficacia esterna del titolo di dottorato di ricerca è influenzata da una serie di variabili, alcune delle quali sono direttamente misurabili, con o senza errore, altre non sono misurabili. Tra le variabili direttamente misurabili si possono senz'altro considerare le caratteristiche anagrafiche, quali l'età ed il genere, e, più in generale, anche il corso di dottorato frequentato, il livello di retribuzione percepito, la professione, la motivazione dell'iscrizione al corso di dottorato, ed altre variabili ancora.

Ai fini della ricerca, gli aspetti più interessanti sono rappresentati dalle variabili non misurabili o latenti. Questo perché si ritiene che gli effetti maggiori sull'”Efficacia esterna” possano derivare da elementi complessi, non misurabili direttamente ed in grado di rappresentare concetti multidimensionali. Tra questi rientrano aspetti come il livello di soddisfazione per l'attività lavorativa svolta, le prospettive e le aspettative per la propria carriera, il grado di coinvolgimento nel proprio lavoro, l'ambiente di lavoro e

forse altri ancora. In particolare, per quanto riguarda gli obiettivi della ricerca, saranno considerate le variabili “Soddisfazione rispetto al lavoro svolto”, “Aspettative” e “Partecipazione lavorativa”, questa ultima intesa come misura del coinvolgimento nell’attività svolta. Tali variabili saranno definite in modo analogo a quanto fatto per l’”Efficacia esterna”, ricorrendo ad indicatori osservati che si assumono tra loro correlati a causa dell’esistenza delle variabili non osservabili o latenti.

Lo strumento di analisi utilizzato è un modello statistico la cui validità sarà verificata utilizzando i dati dell’indagine “Condizione attuale e prospettive occupazionali dei dottori di ricerca”, commissionata dal CNVSU al Dipartimento di Statistica “G. Parenti” dell’Università degli Studi di Firenze. L’analisi di questi aspetti si può fare ricorrendo a modelli che consentono di analizzare un sistema complesso di relazioni e di modellare variabili non direttamente misurabili definite attraverso indicatori osservati che non sono in grado di rappresentare in maniera completamente adeguata gli aspetti considerati, tenendo conto dell’errore di misura. Per questo motivo si ritiene di ricorrere ai modelli ad equazioni strutturali con variabili latenti.

### **1.3 Il modello proposto**

I modelli ad equazioni strutturali (SEMs) costituiscono una metodologia statistica che consente di stimare e verificare ipotesi complesse sulle relazioni di dipendenza tra variabili (esogene ed endogene) osservate o latenti attraverso un sistema di equazioni.

Le relazioni ipotizzate dal modello sono descritte e sintetizzate da un numero limitato di parametri che rappresentano la struttura di covarianza derivata da un insieme di variabili osservate. I modelli, il sistema di equazioni ed i parametri che descrivono le relazioni si dicono “strutturali” perché si riferiscono al sistema complesso di relazioni ipotizzate di cui si vuole studiare la struttura d’insieme. I parametri strutturali misurano l’entità degli effetti, diretti o indiretti, delle variabili esogene (osservate o latenti) sulle variabili endogene. L’effetto totale di una variabile (esogena o endogena) su una variabile endogena è quindi costituito dalla somma degli effetti diretti, che si realizzano tra due variabili senza la mediazione di altre variabili, ed effetti indiretti, che si realizzano quando la relazione tra due variabili è mediata da una o più variabili, chiamate intervenienti.

In presenza di variabili latenti, i modelli ad equazioni strutturali sono composti dal modello strutturale che descrive in termini formali le relazioni ipotizzate tra le variabili latenti attraverso un sistema di equazioni lineari, e dal modello di misura che formalizza le relazioni tra le variabili latenti e le variabili risposta osservate ricorrendo ad un modello fattoriale confermativo (Jöreskog, 1973)<sup>6</sup>. La possibilità di considerare variabili latenti consente di analizzare le relazioni tra variabili che non possono essere misurate perfettamente, o perché non osservabili direttamente, e quindi misurate da una serie di indicatori osservati non completamente adeguati, o perché affette da errore di misura. Tradizionalmente i modelli ad equazioni strutturali hanno considerato solo variabili latenti continue<sup>7</sup> ma sviluppi recenti consentono di estendere l'analisi alle variabili latenti categoriche o a combinazioni tra variabili latenti continue e categoriche con quelli che sono stati chiamati modelli ad equazioni strutturali di seconda generazione (Muthén, 2004). Il modello di misura, inizialmente previsto solo per variabili risposta continue, è stato esteso a variabili categoriche ordinali, dicotomiche, censurate o a combinazioni tra queste, ipotizzando l'esistenza di variabili latenti continue a queste sottostanti (Muthén, 1984). Nella formulazione più generale, il modello di misura è un modello lineare generalizzato e consente di considerare variabili risposta continue, categoriche ordinali e non ordinali, conteggi, durate (a tempo discreto o continuo), comparative o combinazioni tra queste (Bartholomew e Knott, 1999; Rabe-Hasketh e Skrondal, 2004).

In assenza di variabili latenti, potendosi trascurare la componente di misura, il modello complessivo si semplifica e si riduce ad un modello di equazioni simultanee (nella tradizione econometria) o "path analysis" (nella tradizione statistica), il che equivale ad ipotizzare che le variabili siano osservabili, misurate senza errore o con errore trascurabile.

Il modello complessivo può essere rappresentato in forma grafica, e quindi semplificata, con il "path diagram" che descrive il sistema di equazioni che definisce il modello complessivo. Gli sviluppi recenti avvenuti nell'ambito dei modelli grafici consentono di affrontare in termini puramente grafici anche i problemi di identificazione del modello strutturale e di individuazione di modelli equivalenti (Pearl,

---

<sup>6</sup> Recentemente, Asparouhov e Muthén (2009) hanno proposto un modello di misura basato sull'analisi fattoriale esplorativa.

<sup>7</sup> Nel seguito si farà riferimento esclusivo ai modelli ad equazioni strutturali con variabili latenti continue.

2000).

I modelli ad equazioni strutturali hanno dominato l'analisi causale nelle scienze sociali a partire dagli anni cinquanta del secolo scorso. Tuttavia, l'interpretazione in termini causali delle relazioni ipotizzate dal modello deve essere valutata con estrema cautela nel contesto dei disegni osservazionali (Guttman, 1977; Cliff, 1983; Freedman, 1985; 1986; 1992; Holland, 1988; Sobel, 1995). Pearl (2000 e 2009) ha affrontato il problema dell'analisi causale nei modelli ad equazioni strutturali facendo riferimento ai modelli grafici.

I modelli ad equazioni strutturali sono utilizzati per verificare teorie più che per svilupparne di nuove, hanno quindi una natura confermativa più che esplorativa. In questo senso, una volta che una teoria è stata proposta, i modelli possono essere utilizzati per testare la teoria contro i dati osservati, anche se spesso le ipotesi iniziali richiedono qualche forma di adattamento alla luce dell'evidenza empirica.

L'analisi statistica tradizionale assume unità osservate indipendenti e identicamente distribuite. Tale assunzione non è rispettata quando i dati presentano una struttura gerarchica, in cui le unità elementari sono comprese in unità di livello superiore o, più in generale, quando i dati presentano una struttura multilivello<sup>8</sup>. In questo caso, la popolazione osservata risulta costituita da un numero finito di sottopopolazioni tra loro eterogenee composte da unità omogenee per caratteristiche. La presenza di tale struttura, una volta esplicitata, non può essere ignorata poiché, se trascurata, comporta errori standard dei parametri sottostimati e quindi risultati inattendibili dei test di ipotesi inferenziali (Hox, 2002). La soluzione a questo problema dipende dalla natura delle unità di livello superiore e quindi delle sottopopolazioni. Quando l'appartenenza delle unità statistiche elementari alle unità di livello superiore può essere stabilita in maniera attendibile in base ad un criterio osservabile si ricorre ad un'analisi multilivello. Invece, quando tale criterio non è osservabile le sottopopolazioni sono denominate classi latenti e si possono utilizzare i modelli ad equazioni strutturali mistura.

L'approccio tradizionale per estendere i modelli ad equazioni strutturali al caso multilivello quando le variabili risposta osservate sono continue o categoriche modellate da variabili continue latenti sottostanti è quello di formulare modelli separati per le

---

<sup>8</sup> La differenza tra struttura gerarchica e struttura multilivello sarà descritta nel capitolo 2. Per adesso è sufficiente sapere che una struttura dati multilivello comprende anche una struttura gerarchica.

matrici di covarianza entro i gruppi e tra i gruppi. Tale approccio è riconducibile a Goldstein e McDonald (1988) e McDonald e Goldstein (1989), con importanti contributi, tra gli altri, di Muthén (1989, 1994), Lee (1990), Raudenbush (1995), Lee e Poon (1998), Lee e Shi, 2001, Bentler e Liang (2003), Lee e Song, 2004, du Toit e du Toit, 2004 e Liang e Bentler (2004). Quasi tutti questi modelli si basano su stimatori di massima verosimiglianza. Asparouhov e Muthén (2007) hanno proposto un modello ad equazioni strutturali a due livelli che si differenzia dai precedenti perché si basa sul metodo di stima ad informazione limitata e sulla funzione di adattamento dei minimi quadrati ponderati (Muthén, 1984; Muthén, du Toit, e Spisic, 1997).

Rabe-Hesketh, Skrondal e Pickles (2004) hanno proposto un approccio per l'analisi dei modelli ad equazioni strutturali generalizzati multilivello che non prevede la stima di matrici di covarianza separate tra i gruppi ed entro i gruppi ed è basato su uno stimatore di massima verosimiglianza. Tale approccio rientra nell'ambito dei modelli lineari generalizzati misti a variabili latenti e permette di modellare un'ampia gamma di variabili risposta.



# Capitolo 2

## Modelli ad equazioni strutturali multilivello con variabili latenti

### *Introduzione*

L'obiettivo di questo capitolo è quello di descrivere gli aspetti metodologici che caratterizzano i modelli ad equazioni strutturali multilivello con variabili latenti. In particolare, sono trattati la specificazione, la stima e la valutazione dell'adattamento del modello ai dati, prestando attenzione anche agli effetti che possono derivare dal mancato rispetto delle assunzioni previste. Non viene trattato il problema dell'identificazione del modello per il quale si richiederebbe un capitolo specifico. L'organizzazione del capitolo è la seguente. La prima parte riguarda i modelli tradizionali ad un livello (paragrafi 2.1-2.6). Successivamente, dopo aver presentato le caratteristiche principali dell'analisi multilivello (paragrafo 2.7) sono descritti gli aspetti qualificanti dei modelli ad equazioni strutturali a due livelli (paragrafo 2.8) ed i modelli ad equazioni strutturali generalizzati multilivello (paragrafo 2.9).

### **2.1 Specificazione del modello e assunzioni**

I modelli ad equazioni strutturali con variabili latenti sono composti da un modello per le variabili latenti e da un modello di misura. Il modello a variabili latenti è un sistema di equazioni lineari in cui sono specificate le relazioni ipotizzate tra le variabili latenti

esogene ed endogene. Il modello di misura definisce le relazioni tra le variabili latenti e le variabili risposta osservate (indicatori) ed è necessario quando si assume che le variabili osservate siano misure imperfette di variabili latenti sottostanti o siano variabili affette da errore di misura. I modelli ad equazioni strutturali sono stati rappresentati in una varietà di notazioni<sup>9</sup> ma la più diffusa è quella LISREL<sup>10</sup>, a cui si farà riferimento inizialmente (Keesling, 1972; Jöreskog, 1973, 1977; Wiley, 1973).

### 2.1.1 Il modello a variabili latenti

Il modello LISREL assume variabili latenti e variabili risposta osservate continue e distribuite normalmente<sup>11</sup>. Le variabili latenti sono suddivise in endogene, indicate con  $\eta$ , ed esogene, indicate con  $\xi$ .

Il modello lineare per le variabili latenti<sup>12</sup>, riferito all'unità  $i$ , assume la forma seguente:

$$\eta_i = \alpha + B\eta_i + \Gamma\xi_i + \zeta_i,$$

dove  $\eta' = (\eta_1, \eta_2, \dots, \eta_m)$  ( $\eta$ ) è un vettore  $1 \times m$  di variabili casuali latenti endogene<sup>13</sup>,  $\alpha' = (\alpha_1, \alpha_2, \dots, \alpha_m)$  ( $\alpha$ ) è un vettore  $1 \times m$  di intercette,  $\xi' = (\xi_1, \xi_2, \dots, \xi_n)$ <sup>14</sup> ( $\xi$ ) è un vettore  $1 \times n$  di variabili casuali latenti esogene<sup>15</sup> mentre  $\zeta' = (\zeta_1, \zeta_2, \dots, \zeta_m)$ <sup>16</sup> ( $\zeta$ ) è

<sup>9</sup> Una rassegna delle formulazioni alternative si può trovare in Bollen, 1989 (pp. 395-400). Tra queste si ricorda la formulazione RAM (Reticular Action Model) (McArdle, 1980; McArdle e McDonald, 1984) ed il modello Bentler-Weeks (Bentler e Weeks, 1980).

<sup>10</sup> I modelli ad equazioni strutturali sono stati conosciuti inizialmente come modelli JKW (Bentler, 1980), acronimo derivato dalle iniziali dei tre autori Jöreskog, Keesling e Wiley ma hanno acquistato popolarità con il nome LISREL (Linear Structural Relationship) che deriva da quello del programma software statistico sviluppato nella sua prima versione nel 1973 da Karl Jöreskog e Dag Sörbom per stimare i parametri dei modelli ad equazioni strutturali con il metodo di massima verosimiglianza.

<sup>11</sup> L'assunzione di normalità delle variabili osservate sarà rimossa in seguito.

<sup>12</sup> Il modello per le variabili latenti viene comunemente definito anche "modello strutturale". Tuttavia, questa definizione è stata ritenuta ingannevole perché considera "strutturali" solo le equazioni che compongono il modello a variabili latenti. Secondo una visione più ampia, tutte le equazioni, quelle del modello a variabili latenti e quelle del modello di misura, descrivono relazioni strutturali (Bollen, 1989).

<sup>13</sup> Le variabili endogene sono quelle variabili influenzate da altre variabili del modello ma non dai termini di disturbo.

<sup>14</sup> Il vettore  $\xi$  è, nella maggior parte dei casi, un vettore di variabili casuali ma più avanti tale assunzione sarà rimossa (Muthén, 1983, 1984).

<sup>15</sup> Le variabili esogene sono quelle variabili non spiegate dal modello ed incorrelate con i termini di disturbo.

<sup>16</sup> Il vettore  $\zeta$  è costituito, di solito, da variabili casuali, ad eccezione del caso in cui un'equazione

un vettore  $1 \times m$  di errori o termini di disturbo casuali, ciascuno dei quali comprende quelle variabili non incluse nel modello che influenzano le corrispondenti variabili latenti endogene.

Le relazioni tra le variabili latenti sono rappresentate dalla matrice  $\mathbf{B}$  di ordine  $m \times m$  i cui elementi  $(\beta_{ij})$  sono i coefficienti che legano le variabili endogene  $(\eta_i)$  tra di loro, e dalla matrice  $\mathbf{\Gamma}$  di ordine  $m \times n$ , i cui elementi  $\gamma_{ij}$  sono i coefficienti che legano le variabili endogene  $(\eta_i)$  con le variabili esogene  $(\xi_i)$ . La diagonale principale della matrice  $\mathbf{B}$  contiene i coefficienti che legano ogni variabile endogena con se stessa  $(\beta_{ii})$  ed è sempre costituita da elementi nulli. In questo modo si esclude la possibilità che una variabile endogena abbia effetto su se stessa. Gli elementi nulli al di fuori della diagonale principale indicano invece l'assenza di effetti tra le variabili endogene. In generale, gli elementi nulli e non nulli delle matrici  $\mathbf{B}$  e  $\mathbf{\Gamma}$  sono imposti considerando la teoria di fondo alla costruzione del modello.

Il modello a variabili latenti è definito anche da due matrici di covarianza, la matrice di covarianza tra le variabili latenti esogene e la matrice di covarianza degli errori. La matrice di covarianza tra le variabili latenti esogene è  $\Phi$  (phi), di ordine  $n \times n$  i cui elementi  $(\phi_{ij})$  consentono di specificare le relazioni ipotizzate tra le variabili esogene incluse nel modello. La matrice di covarianza degli errori è  $\Psi$  (psi), di ordine  $m \times m$ , i cui elementi  $(\psi_{ij})$  consentono di considerare l'effetto sulle variabili endogene di variabili non previste dal modello. In particolare, gli elementi sulla diagonale principale  $(\psi_{ii})$  rappresentano la varianza delle variabili endogene  $\eta_i$  non spiegata dalle variabili esplicative incluse nelle equazioni<sup>17</sup> mentre gli elementi al di fuori della diagonale principale  $(\psi_{ij})$  rappresentano gli effetti di variabili non previste che agiscono contemporaneamente su due variabili endogene. La presenza della matrice  $\Psi$  (psi) è quindi necessaria per evitare il rischio di osservare una correlazione spuria (e quindi inesistente) tra due variabili endogene che si verifica quando tra i fattori omessi riassunti in  $\zeta_i$  ci sono variabili che agiscono contemporaneamente su entrambe le variabili endogene. In questo caso, prevedere una correlazione tra gli errori, riassunta

---

rappresenti un'identità  $\eta_i = \eta_2 + \eta_3$ , in cui  $\zeta_1 = 0$  e quindi costante (Bollen, 1989 pp. 13).

<sup>17</sup> In un modello specificato in maniera corretta, cioè che include tutte le variabili che hanno effetto sulle variabili endogene, il termine di errore di ciascuna equazione rappresenta una grandezza trascurabile.

nel parametro  $\psi_{ij}$ , fa scomparire la correlazione spuria ed equivale ad assumere l'esistenza di una variabile esterna al modello che ha effetto su entrambe le variabili endogene.

La specificazione degli elementi nulli e non nulli delle matrici  $\mathbf{B}$  e  $\mathbf{\Psi}$  consente di distinguere i modelli *ricorsivi* dai modelli *non ricorsivi*. I modelli ricorsivi non prevedono effetti reciproci tra le variabili latenti endogene e questo comporta una matrice  $\mathbf{B}$  triangolare inferiore e una matrice di covarianza degli errori,  $\mathbf{\Psi}$ , diagonale<sup>18</sup>, in quanto gli errori tra le diverse equazioni non sono correlati. I modelli non ricorsivi prevedono invece relazioni reciproche tra le variabili latenti endogene, quindi gli elementi della matrice  $\mathbf{B}$  sono parametri incogniti, compresi quelli della parte triangolare superiore. Inoltre, la matrice  $\mathbf{\Psi}$  non è più diagonale e contiene elementi non nulli anche al di fuori della diagonale principale.

Il modello per le variabili latenti assume che  $\boldsymbol{\eta}_i \sim N_n(\mathbf{0}; \mathbf{I})$ ,  $\boldsymbol{\xi}_i \sim N_n(\mathbf{0}; \boldsymbol{\Phi})$ ,  $\boldsymbol{\zeta}_i \sim N_m(\mathbf{0}; \boldsymbol{\Psi})$  e, per evitare stime inconsistenti dei coefficienti, che i fattori omessi riassunti in  $\boldsymbol{\zeta}_i$  non siano correlati con le variabili esogene  $\boldsymbol{\xi}_i$ , cioè  $Cov(\boldsymbol{\xi}_i', \boldsymbol{\zeta}_i) = 0$ . Inoltre, assume che i termini di errore  $\zeta_i$  siano omoschedastici e non autocorrelati, dove l'omoschedasticità implica la varianza costante tra le osservazioni, cioè  $E(\zeta_{ij}^2) = Var(\zeta_i)$  per ogni osservazione  $j$  mentre l'autocorrelazione implica assenza di correlazione tra le osservazioni, cioè  $Cov(\zeta_{ij}, \zeta_{il}) = 0$  per  $j \neq l$ . Infine, si assume che  $(\mathbf{I} - \mathbf{B})$  sia una matrice non singolare<sup>19</sup> tale che esista la sua inversa  $(\mathbf{I} - \mathbf{B})^{-1}$ , condizione necessaria affinché il modello per le variabili latenti possa essere scritto in forma ridotta<sup>20</sup>.

I parametri delle matrici  $\mathbf{B}$ ,  $\boldsymbol{\Gamma}$  e  $\mathbf{\Psi}$  possono essere incogniti, fissi, e vincolati. I parametri liberi sono quelli incogniti da stimare in base ai dati osservati e riassumono le relazioni specificate dal modello. I parametri fissi hanno valori stabiliti a priori e, a differenza dei parametri incogniti, sono mantenuti costanti durante la procedura di

<sup>18</sup> Se  $\mathbf{\Psi}$  non è diagonale ma le altre condizioni sono rispettate, il modello si dice *parzialmente ricorsivo*. Occasionalmente il termine ricorsivo viene utilizzato per riferirsi a qualunque modello con una matrice  $\mathbf{B}$  triangolare inferiore, indipendentemente dal fatto che  $\mathbf{\Psi}$  sia diagonale (Bollen, 1989).

<sup>19</sup> Per questo è necessario che nessuna variabile endogene si possa esprimere come combinazione lineare di altre variabili endogene  $\eta_i$ , il che significa che le equazioni del modello devono essere indipendenti tra loro.

<sup>20</sup> La forma ridotta prevede le variabili endogene da un lato dell'equazione e le variabili esogene dall'altro.

stima. Il loro valore viene di solito fissato a zero per rappresentare l'assenza di relazione tra le variabili corrispondenti ma è possibile fissare un valore diverso, qualora la teoria suggerisca l'esistenza di una relazione significativa. I parametri vincolati hanno valori imposti uguali ad altri parametri ma, a differenza dei parametri fissi, il loro valore non è stabilito in anticipo. I parametri vincolati sono inclusi in un modello quando la restrizione imposta è derivabile dalla teoria sottostante o rappresenta una ipotesi interessante da verificare.

### 2.1.2 Il modello di misura

Il modello di misura definisce la relazione tra le variabili latenti (endogene ed esogene) e le variabili risposta osservate con un modello di analisi fattoriale confermativa (Jöreskog, 1969).

Le variabili osservate si distinguono a seconda che siano riferite alle variabili latenti endogene, indicate con  $y$ , o alle variabili latenti esogene, indicate con  $x$ . Le variabili osservate si assumono continue e distribuite normalmente,  $\mathbf{y} \sim N_p(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})$  e  $\mathbf{x} \sim N_q(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$ .

Il modello di misura è composto da due equazioni, una per le variabili endogene e l'altra per le variabili esogene:

$$\begin{aligned}\mathbf{y}_i &= \mathbf{v}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \\ \mathbf{x}_i &= \mathbf{v}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi}_i + \boldsymbol{\delta}_i,\end{aligned}$$

dove  $\mathbf{y}' = (y_1, y_2, \dots, y_p)$  è un vettore  $1 \times p$  di indicatori delle variabili latenti endogene contenute in  $\boldsymbol{\eta}$ ,  $\mathbf{v}_y' = (v_{y_1}, v_{y_2}, \dots, v_{y_p})$  è un vettore  $1 \times p$  di intercette,  $\boldsymbol{\Lambda}_y$  è una matrice  $p \times m$  di coefficienti fattoriali  $(\lambda_{ij}^y)$  che legano le variabili latenti endogene ai rispettivi indicatori e  $\boldsymbol{\varepsilon}' = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)$  è un vettore  $1 \times p$  di errori di misura. Analogamente,  $\mathbf{x}' = (x_1, x_2, \dots, x_q)$  è un vettore  $1 \times q$  di indicatori delle variabili latenti esogene contenute in  $\boldsymbol{\xi}$ ,  $\mathbf{v}_x' = (v_{x_1}, v_{x_2}, \dots, v_{x_q})$  è un vettore  $1 \times q$  di intercette,  $\boldsymbol{\Lambda}_x$  è una matrice  $q \times n$  di coefficienti fattoriali  $(\lambda_{ij}^x)$  che legano le variabili latenti esogene ai rispettivi

indicatori e  $\delta' = (\delta_1, \delta_2, \dots, \delta_q)$  è un vettore  $1 \times q$  di errori di misura.

La presenza dei termini di errore, rispettivamente  $\varepsilon_i$  e  $\delta_i$ , consente di tenere in considerazione, per ciascuna variabile osservata, la variabilità non spiegata, dovuta all'esistenza di fattori non osservati perché misurati da indicatori imperfetti o all'errore di misura.

Anche il modello di misura prevede due matrici di covarianza, la matrice di covarianza  $\Theta_\varepsilon$  di ordine  $p \times p$  degli errori di misura  $(\varepsilon_{ij})$  riferiti agli indicatori contenuti nel vettore  $y_i$  e la matrice di covarianza  $\Theta_\delta$  di ordine  $q \times q$  degli errori di misura  $(\delta_{ij})$  riferiti agli indicatori contenuti nel vettore  $x_i$ . Gli elementi sulla diagonale principale di ognuna di queste matrici rappresentano la varianza di ciascuna variabile osservata (indicatore) non spiegata dalle variabili latenti mentre gli elementi al di fuori della diagonale principale rappresentano la covarianza tra gli errori di misura associati ai diversi indicatori. La presenza di elementi al di fuori della diagonale principale è quindi indicativa dell'esistenza di correlazione tra gli errori.

Il modello di misura assume che i termini di errore<sup>21</sup> abbiano media nulla e non siano correlati con le variabili latenti:  $E(\varepsilon) = 0$ ,  $E(\delta) = 0$ ,  $Cov(\eta, \varepsilon') = 0$  e  $Cov(\xi, \delta') = 0$ <sup>22</sup>. Inoltre, assume che per una data equazione gli errori  $\varepsilon_i$  e  $\delta_i$  siano omoschedastici e non autocorrelati anche se la varianza degli errori può assumere valori diversi per ciascuna equazione del modello di misura.

In definitiva, per stimare il modello generale, è necessario specificare gli elementi nelle quattro matrici dei coefficienti strutturali,  $B$ ,  $\Gamma$ ,  $\Lambda_y$ ,  $\Lambda_x$  e nelle quattro matrici di covarianza,  $\Phi$ ,  $\Psi$ ,  $\Theta_\varepsilon$ ,  $\Theta_\delta$ . Nella fase di specificazione del modello, quando è necessario stabilire quali variabili latenti endogene sono legate tra di loro, la relazione tra le variabili latenti, gli indicatori nel modello di misura e l'eventuale esistenza di correlazione tra gli errori di equazioni diverse, diventa fondamentale la conoscenza dell'ambito di ricerca in modo da evitare un modello sotto specificato.

Asparouhov e Muthén (2009) hanno proposto un approccio ai modelli ad

---

<sup>21</sup> Nell'analisi fattoriale gli errori  $\delta_i$  e  $\varepsilon_i$  si dicono anche fattori unici e sono scomposti in una parte specifica, propria di ciascuna variabile, ed una non specifica o casuale.

<sup>22</sup> Una correlazione tra gli errori di misura  $\delta_i$  e  $\varepsilon_i$  e le rispettive variabili latenti,  $\eta_i$  e  $\xi_i$  (endogeneità), può portare a stimatori inconsistenti dei parametri.

equazioni strutturali in cui il modello di misura è basato sull'analisi fattoriale esplorativa (EFA), in aggiunta o in sostituzione al tradizionale modello di analisi fattoriale confermativa (CFA). Tale approccio descrive un modello sperimentale ad equazioni strutturali (ESEM) che diventa importante nelle situazioni di incertezza sulla capacità rappresentativa del modello di misura perché consente di estendere i modelli ad equazioni strutturali tradizionali a modelli di misurazione meno restrittivi, permettendo così un insieme più ricco di alternative rispetto al modello a priori.

Il contributo di Asparouhov e Muthén si fonda sulla considerazione che un modello di misura basato sull'analisi fattoriale confermativa, consente di specificare a priori i fattori che influenzano le variabili osservate (indicatori), i cui coefficienti fattoriali rappresentano i parametri incogniti, e di fissare a zero i coefficienti dei fattori che si assume non abbiano effetto<sup>23</sup>. In questo modo la conoscenza dell'ambito di ricerca svolge un ruolo decisivo non solo perché rende la definizione delle variabili latenti più fondata dal punto di vista teorico ma anche e soprattutto perché viene incorporata nel modello sotto forma di restrizioni che consentono di formalizzare modelli di misura parsimoniosi (semplificati). Tuttavia, se l'utilizzo del modello fattoriale confermativo è attraente per i motivi sopra esposti, la possibilità di stabilire a piacere i parametri incogniti da stimare può portare ad avere troppi coefficienti fattoriali uguali a zero e quindi a specificare modelli di misura più parsimoniosi del dovuto<sup>24</sup>. Per questo motivo, accade spesso che un modello non si adatti bene ai dati osservati e deve essere modificato<sup>25</sup>. Inoltre, quando un modello di misura è mal specificato, le correlazioni tra i fattori con coefficienti diversi da zero tendono ad essere sovrastimate e, di conseguenza, le relazioni tra le variabili latenti sono distorte. In situazioni in cui non si ha certezza sulla capacità rappresentativa del modello di misura è quindi preferibile utilizzare un modello fattoriale esplorativo, soprattutto se si considera che la procedura di modifica e miglioramento di un modello fattoriale confermativo implica

---

<sup>23</sup> Spesso si ipotizza una struttura fattoriale molto semplice, in cui ciascun indicatore è influenzato da un solo fattore, in modo da evitare coefficienti fattoriali incrociati (quelli che si hanno quando un indicatore è influenzato da più di un fattore).

<sup>24</sup> Infatti, il numero di restrizioni nei modelli fattoriali confermativi è tipicamente molto più grande rispetto al numero di restrizioni richieste per identificare il modello di misura che, in un'analisi fattoriale esplorativa con  $m$  fattori, è pari a  $m^2$  restrizioni imposte sui parametri (coefficienti, varianze e covarianze fattoriali). Di conseguenza, spesso ci sono troppi coefficienti incrociati con valore nullo, molti di più di quanti siano giustificati dalla teoria.

<sup>25</sup> Per la ricerca del modello più adatto, di solito ci si avvale degli indici di modifica del modello, il cui utilizzo è stato peraltro criticato (MacCallum, Roznowski e Necowitz, 1992).

un'analisi esplorativa (Browne, 2001).

I modelli ad equazioni strutturali basati sull'analisi fattoriale esplorativa risultano quindi meno restrittivi e possono affiancare i tradizionali modelli di misura basati sull'analisi fattoriale confermativa aggiungendo flessibilità al modello di misura quando il modello fattoriale esplorativo risulta più aderente alla realtà osservata, per la scarsa conoscenza teorica delle relazioni tra variabili osservate e variabili latenti o per la sua maggiore complessità rispetto a quella prevista dall'analisi confermativa. Il modello fattoriale confermativo rimane invece preferibile in tutte le situazioni in cui il modello di misura è ben definito dal punto di vista teorico.

### *2.1.3 Il modello con variabili esogene osservate*

Il modello assume una forma particolare in presenza di variabili esplicative esogene osservate, misurate senza errore o con errore trascurabile (Muthén, 1983, 1984). In questo caso le variabili osservate coincidono con le variabili latenti esogene ed il modello di misura per le variabili latenti esogene si semplifica, in quanto  $\mathbf{x}_i = \boldsymbol{\xi}_i$ . Il modello per le variabili latenti assume invece la forma seguente:

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i,$$

dove  $\mathbf{x}_i' = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_q)$  è un vettore  $1 \times q$  di variabili esplicative osservate. In questo caso, rispetto al modello convenzionale,  $\mathbf{B}$  è una matrice di ordine  $m \times m$  di coefficienti che legano le variabili latenti tra di loro e  $\boldsymbol{\Gamma}$  è una matrice di ordine  $q \times m$  di coefficienti che legano le variabili latenti con le variabili esplicative osservate.

Il modello specificato da Muthén è condizionato alle variabili esplicative  $\mathbf{x}_i$  e, a differenza del modello a variabili latenti tradizionale, in cui le variabili osservate sono solo variabili risposta (o dipendenti), non richiede assunzioni particolari sulla distribuzione delle variabili esplicative esogene.

### *2.1.4 La scomposizione degli effetti tra variabili*

I modelli ad equazioni strutturali consentono di scomporre gli effetti di una variabile su un'altra, in effetti diretti, indiretti e totali. Gli effetti diretti sono quelli che si realizzano



tra due variabili senza la mediazione di altre variabili e sono compresi nelle matrici strutturali  $\mathbf{B}$ ,  $\mathbf{\Gamma}$ ,  $\mathbf{\Lambda}_y$ ,  $\mathbf{\Lambda}_x$ . Gli effetti indiretti si realizzano invece quando la relazione tra due variabili è mediata da almeno un'altra variabile, denominata interveniente. Gli effetti totali sono costituiti dalla somma degli effetti diretti ed indiretti. In particolare, gli effetti totali sono stati definiti in due modi alternativi che producono il medesimo risultato. Il primo fa riferimento alla somma delle potenze delle matrici dei coefficienti mentre il secondo fa ricorso alla forma ridotta dei coefficienti<sup>26</sup>.

Il metodo della somma dei coefficienti delle matrici definisce gli effetti totali tra due variabili latenti  $\eta$  e  $\eta$  nel modo seguente (Fox, 1980):

$$\mathbf{T}_{\eta\eta} = \sum_{k=1}^{\infty} \mathbf{B}^k .$$

$\mathbf{T}_{\eta\eta}$  è definito solo se la sommatoria converge ad una matrice con un numero finito di elementi.

Gli effetti indiretti si ottengono invece per differenza tra gli effetti totali e gli effetti diretti. In termini generali, per i modelli ricorsivi<sup>27</sup> in cui  $\mathbf{B}$  è una matrice triangolare inferiore,  $\mathbf{B}^k$  è uguale a zero per  $k \geq m$ , dove  $m$  è il numero di variabili latenti  $\eta$ . Perciò,  $\mathbf{B}$  converge e gli effetti totali sono definiti. Gli effetti sono quindi uguali a  $\mathbf{T}_{\eta\eta} = \sum_{k=1}^{m-1} \mathbf{B}^k$  mentre gli effetti indiretti sono dati da:

$$\mathbf{I}_{\eta\eta} = \mathbf{T}_{\eta\eta} - \mathbf{B} = \sum_{k=2}^{m-1} \mathbf{B}^k .$$

In generale, affinché l'effetto totale  $\mathbf{T}_{\eta\eta}$  sia definito, la matrice  $\mathbf{B}^k$  deve convergere a zero quando  $k \rightarrow \infty$  (Ben-Israel e Greville, 1974). La matrice è convergente se e solo se il valore assoluto dell'autovalore maggiore è minore di uno (Bentler e Freeman, 1983). In questo caso, l'effetto totale si può scrivere nel modo seguente:

---

<sup>26</sup> Per semplicità la descrizione è limitata al primo metodo.

<sup>27</sup> L'analisi è limitata ai modelli ricorsivi, per i modelli non ricorsivi si può fare riferimento a Bollen (1987).

$$\mathbf{T}_{\eta\eta} = (\mathbf{B} - \mathbf{I})^{-1} - \mathbf{I},$$

mentre gli effetti indiretti si ricavano sottraendo  $\mathbf{B}$  a  $\mathbf{T}_{\eta\eta}$ :

$$\mathbf{I}_{\eta\eta} = (\mathbf{B} - \mathbf{I})^{-1} - \mathbf{I} - \mathbf{B}.$$

La scomposizione degli effetti delle variabili latenti esogene sulle variabili latenti endogene si ottiene in maniera analoga:

$$\mathbf{T}_{\eta\xi} = (\mathbf{B} - \mathbf{I})^{-1} \mathbf{\Gamma}.$$

Poiché gli effetti di  $\xi$  su  $\eta$  sono compresi in  $\mathbf{\Gamma}$ , gli effetti indiretti di  $\xi$  su  $\eta$  sono:

$$\mathbf{I}_{\eta\xi} = ((\mathbf{B} - \mathbf{I})^{-1} \mathbf{I}) \mathbf{\Gamma}.$$

Tali effetti sono quindi uguali al prodotto degli effetti totali tra le variabili endogene  $\eta$  moltiplicati per gli effetti diretti tra le variabili esogene  $\xi$  sulle endogene  $\eta$ .

## 2.2 Le matrici implicate dei momenti

Una volta specificato il modello, è possibile esprimere i momenti di primo e secondo ordine (medie, varianze e covarianze) delle variabili osservate nella popolazione in funzione dei parametri del modello, riassunti per comodità nel vettore  $\boldsymbol{\theta} = (\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Lambda}_y, \mathbf{\Lambda}_x, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Theta}_\varepsilon, \mathbf{\Theta}_\delta)$ . Se il modello è specificato in maniera corretta, il vettore delle medie e la matrice di covarianza della popolazione sono riprodotti perfettamente in funzione dei parametri incogniti e sussistono le seguenti relazioni:

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}),$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}).$$

dove  $\boldsymbol{\mu}$  (mu) e  $\boldsymbol{\Sigma}$  (sigma) sono rispettivamente il vettore delle medie e la matrice di covarianza delle variabili osservate nella popolazione e  $\boldsymbol{\theta}$  (theta) è il vettore dei parametri incogniti del modello.

Il modello specificato, espresso in funzione delle relazioni ipotizzate e riassunte dai parametri contenuti nelle otto matrici, implica quindi un vettore delle medie e una matrice di covarianza tra le variabili osservate, rispettivamente il vettore delle medie e la matrice di covarianza implicati attesi. Il vettore dei parametri  $\boldsymbol{\theta}$  definisce in questo modo la forma di un particolare modello ad equazioni strutturali.

Per ricavare i momenti implicati, è preferibile esprimere il modello in forma ridotta, in cui ciascuna variabile endogena risulta funzione solo delle variabili esogene, dei coefficienti e degli errori. Il modello per le variabili latenti in forma ridotta si ottiene quindi risolvendo l'equazione in modo da esplicitare le variabili endogene:

$$\boldsymbol{\eta}_i = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\zeta}_i).$$

Il modello di misura per le variabili  $\mathbf{y}_i$  in forma ridotta si ottiene sostituendo il modello per le variabili latenti:

$$\mathbf{y}_i = \mathbf{v}_y + \boldsymbol{\Lambda}_y(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\zeta}_i).$$

Il modello di misura per le variabili  $\mathbf{x}_i$  è, invece, già espresso in forma ridotta.

Il vettore delle medie della popolazione implicato in funzione dei parametri del modello assume la forma seguente:

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{v}_y + \boldsymbol{\Lambda}_y(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\alpha} + \boldsymbol{\Gamma}\boldsymbol{\mu}\boldsymbol{\xi}) \\ \mathbf{v}_x + \boldsymbol{\Lambda}_x\boldsymbol{\mu}\boldsymbol{\xi} \end{bmatrix},$$

dove il primo elemento è il vettore implicato delle medie delle variabili  $\mathbf{y}_i$  mentre il secondo è il vettore implicato delle medie delle variabili  $\mathbf{x}_i$ .

In modo analogo, si può esprimere in funzione dei parametri del modello anche la matrice di covarianza implicata per le variabili osservate  $\mathbf{x}_i$  e  $\mathbf{y}_i$  (Jöreskog, 1977):

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{\Sigma}_{yy}(\boldsymbol{\theta}) & \mathbf{\Sigma}_{yx}(\boldsymbol{\theta}) \\ \mathbf{\Sigma}_{xy}(\boldsymbol{\theta}) & \mathbf{\Sigma}_{xx}(\boldsymbol{\theta}) \end{bmatrix},$$

le cui componenti sono, rispettivamente le seguenti:

$$\mathbf{\Sigma}_{yy}(\boldsymbol{\theta}) = \mathbf{\Lambda}_y (\mathbf{I} - \mathbf{B})^{-1} (\mathbf{\Gamma} \mathbf{\Phi} \mathbf{\Gamma}' + \mathbf{\Psi}) (\mathbf{I} - \mathbf{B})^{-1} \mathbf{\Lambda}_y' + \mathbf{\Theta}_\varepsilon,$$

$$\mathbf{\Sigma}_{xx}(\boldsymbol{\theta}) = \mathbf{\Lambda}_x \mathbf{\Phi} \mathbf{\Lambda}_y' + \mathbf{\Theta}_\delta,$$

$$\mathbf{\Sigma}_{xy}(\boldsymbol{\theta}) = \mathbf{\Lambda}_x \mathbf{\Phi} \mathbf{\Gamma}' (\mathbf{I} - \mathbf{B})^{-1} \mathbf{\Lambda}_y'.$$

Le espressioni per i momenti di primo e secondo ordine implicati sono espressi in forma sufficientemente generale e consentono di ottenere la matrice di covarianza implicata per qualunque modello ad equazioni strutturali. In generale, la relazione tra  $\mathbf{\Sigma}$  e  $\mathbf{\Sigma}(\boldsymbol{\theta})$  è essenziale per la comprensione dei problemi di identificazione, di stima e di valutazione di adattamento del modello.

### 2.3 Modelli ad equazioni strutturali con variabili risposta categoriche

I modelli ad equazioni strutturali convenzionali si basano sull'assunzione di continuità e normalità multivariata delle variabili osservate. In pratica, tale assunzione è difficilmente soddisfatta poiché spesso gli indicatori che misurano le variabili latenti sono di natura categorica e presentano un numero finito e limitato di modalità, ordinate o non ordinate.

In presenza di variabili osservate categoriche le conseguenze che derivano non sono trascurabili (Bollen, 1989): i) la relazione tra le variabili osservate e le variabili latenti (modello di misura) non è più lineare,  $\mathbf{y}_i^c \neq \mathbf{v}_y + \mathbf{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i$  e  $\mathbf{x}_i^c \neq \mathbf{v}_x + \mathbf{\Lambda}_x \boldsymbol{\xi}_i + \boldsymbol{\delta}_i$ , dove  $\mathbf{y}_i^c$  e  $\mathbf{x}_i^c$  sono i vettori di variabili osservate categoriche; ii) il modello ipotizzato di

solito<sup>28</sup> non riflette la reale struttura delle relazioni nella popolazione, espressa dalla matrice di covarianza  $\Sigma_0^c \neq \Sigma(\theta)$ , dove  $\Sigma_0^c$  è la matrice di covarianza delle variabili categoriche osservate nella popolazione.

In questo caso, per mantenere la relazione lineare tra le variabili risposta osservate e le variabili latenti è necessario modificare il modello di misura mentre il modello per le variabili latenti rimane invariato. In generale, si possono seguire due approcci per estendere il modello di misura. Il primo assume l'esistenza di una variabile latente sottostante alle variabili osservate e consente di analizzare modelli ad equazioni strutturali con variabili risposta continue, categoriche ordinali, dicotomiche, censurate e combinazioni tra queste (Muthén, 1978, 1983, 1984). Il secondo considera la classe dei modelli lineari generalizzati e consente di analizzare modelli ad equazioni strutturali con una gamma più ampia di variabili risposta, in cui sono comprese anche frequenze, risposte categoriche non ordinali, graduatorie, durate (a tempo continuo o discreto) e combinazioni tra queste (Bartholomew and Knott, 1999; Rabe-Hasketh e Skrondal, 2004)<sup>29</sup>. In entrambi i casi, il modello di misura può essere esteso per includere anche variabili esplicative (Muthén, 1979, 1983, 1984).

### 2.3.1 *Modello di misura con variabili risposta latenti*

L'approccio prevalente nei modelli ad equazioni strutturali quando le variabili risposta sono categoriche dicotomiche, ordinali, censurate o combinazioni tra queste assume che ciascuna variabile osservata sia un indicatore di una variabile risposta non osservabile o latente<sup>30</sup> continua, i cui valori sono osservabili unicamente attraverso le variabili risposta categoriche.

Le variabili latenti sottostanti alle variabili osservate categoriche hanno un campo di variazione compreso tra  $-\infty$  e  $+\infty$  ed assegnano una metrica alle corrispondenti variabili categoriche. Le variabili risposta latenti sono trasformate in variabili categoriche aventi un numero finito di modalità osservabili attraverso un

<sup>28</sup> Si utilizza l'avverbio "di solito" perché è possibile avere una struttura di covarianza valida nel caso di variabili continue e categoriche (Muthén e Kaplan, 1985).

<sup>29</sup> Questo argomento sarà affrontato nel paragrafo 2.9, in cui sono descritti i modelli ad equazioni strutturali generalizzati multilivello.

<sup>30</sup> Queste variabili latenti hanno una natura diversa dalle variabili latenti endogene o esogene definite in precedenza e servono solo per trattare le variabili risposta categoriche, in modo da mantenere la relazione lineare nel modello di misura. Dal punto di vista astratto (teorico) tali variabili latenti rappresentano la "propensione continua non osservabile ad assumere modalità categoriche osservabili".

modello soglia<sup>31</sup> (threshold model). Il modello soglia suddivide la distribuzione delle variabili latenti continue e consente di stabilire una relazione con le variabili osservate categoriche attraverso una serie di parametri, denominati soglie. La formalizzazione del modello ad equazioni strutturali richiede quindi un insieme aggiuntivo di equazioni necessarie a legare gli indicatori latenti continui ai corrispondenti indicatori categorici osservati:

$$\begin{aligned} \mathbf{y}_i^* &= f(\mathbf{y}_i, \boldsymbol{\tau}_y), \\ \mathbf{x}_i^* &= f(\mathbf{x}_i, \boldsymbol{\tau}_x), \end{aligned}$$

dove  $\mathbf{y}_i^*$  e  $\mathbf{x}_i^*$  sono i vettori di variabili latenti continue,  $\mathbf{y}_i$  e  $\mathbf{x}_i$  sono i vettori di variabili osservate continue, categoriche dicotomiche, ordinali o censurate,  $\boldsymbol{\tau}_y$  e  $\boldsymbol{\tau}_x$  sono i rispettivi vettori dei parametri soglia che determinano i valori assunti dai corrispondenti vettori di indicatori osservati e  $f(\cdot)$  è la funzione che lega le variabili latenti continue alle variabili categoriche.

Nel caso di un generico indicatore osservato  $y_i$ , dove  $i = 1, 2, \dots, n$  si riferisce alle unità osservate si definisce la corrispondente variabile latente di risposta continua  $y_i^*$ . Il vantaggio di questa formulazione è che il modello di misura utilizzato per le variabili risposta continue può essere specificato anche per le variabili latenti continue  $y_i^*$  sottostanti alle variabili risposta osservate  $y_i$ .

Quando la variabile osservata  $y$  è continua il modello soglia non è necessario poiché la variabile latente sottostante si osserva direttamente,  $y_i = y_i^*$ . Quando la variabile risposta osservata è categorica con  $C$  modalità ordinate, la variabile di risposta latente si definisce nel modo seguente:

$$y = k \quad \Leftrightarrow \quad \tau_k < y_i^* \leq \tau_{k+1},$$

dove  $k = 0, 1, \dots, K-1$  e  $-\infty = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < \tau_K = +\infty$ . In generale, nel caso

---

<sup>31</sup> In conseguenza, i punti che suddividono le variabili latenti continue in variabili categoriche si chiamano soglie.

di  $K$  modalità i parametri soglia sono  $K - 1$ ,  $\tau_1, \tau_2, \dots, \tau_{K-1}$ . Quando la variabile risposta  $y$  è dicotomica, si può considerare come caso particolare di una variabile categorica ordinale con due sole modalità, in cui  $K = 2$ . In questo caso il parametro soglia è uno solo,  $\tau_1$ , e la variabile latente si definisce nel modo seguente:

$$y = k \quad \Leftrightarrow \quad y_i^* > \tau_1.$$

Quando la variabile osservata  $y$  è censurata dal basso<sup>32</sup>, la variabile latente sottostante si definisce nel modo seguente:

$$\begin{aligned} y_i &= y_i^* & \text{se} & \quad y_i^* \geq L, \\ y_i &= L & \text{se} & \quad y_i^* < L, \end{aligned}$$

dove  $L$  rappresenta il valore al di sotto del quale la variabile non è osservata.

In generale, la stima delle soglie si può fare se si conosce la distribuzione della variabile latente continua  $y^*$  e la proporzione di osservazioni in ciascuna modalità della variabile osservata  $y$ . In linea di principio si può scegliere qualunque funzione di densità continua per la variabile  $y^*$  e, considerato che l'informazione contenuta nei dati è di natura ordinale, la sua distribuzione è determinata a meno di una trasformazione monotona. Tuttavia, poiché qualunque variabile continua avente una funzione di densità e di distribuzione può essere ricondotta ad una funzione di densità normale in seguito ad una trasformazione monotona, è preferibile scegliere la distribuzione normale standardizzata, avente funzione di densità  $\phi(u)$  e funzione di distribuzione  $\Phi(u)$ . L'assunzione più comune è quindi quella che gli indicatori seguano una distribuzione normale multivariata, tale che le distribuzioni marginali delle variabili singolarmente considerate siano normali. In questo caso, la probabilità di osservare una risposta nella modalità  $c$  è data da:

---

<sup>32</sup> Equazioni analoghe si possono definire per indicatori censurati dall'alto o per indicatori censurati dal basso e dall'alto.

$$\pi_i = \Pr(y_i = k) = \Pr(\tau_k < y_i^* \leq \tau_{k+1}) = \int_{\tau_k}^{\tau_{k+1}} \phi(u) du = \Phi(\tau_k) - \Phi(\tau_{k+1}),$$

tale che:

$$\tau_k = \Phi^{-1}(\pi_1 + \pi_2 + \dots + \pi_k), \quad k = 1, 2, \dots, K-1,$$

dove  $\Phi^{-1}(\cdot)$  è l'inversa della funzione di densità normale standardizzata mentre  $(\pi_1 + \pi_2 + \dots + \pi_k)$  è la probabilità di osservare una risposta nella modalità  $c$  o in una modalità precedente. Poiché  $\pi_{ij}$  sono le probabilità incognite della popolazione, in pratica si possono stimare con le corrispondenti frequenze relative associate alle risposte osservate in ciascuna modalità. In tal caso una stima delle soglie è data da:

$$\hat{\tau}_k = \Phi^{-1}(p_1 + p_2 + \dots + p_k), \quad k = 1, 2, \dots, K-1,$$

dove  $(p_1 + p_2 + \dots + p_k)$  è la probabilità campionaria di osservare una risposta nella modalità  $c$  o in una modalità precedente.

Il modello di misura per le variabili risposta latenti si può definire, in maniera generica, nel modo seguente:

$$y_i^* = \nu_i + \varepsilon_i,$$

dove  $\nu_i$  è il predittore lineare mentre  $\varepsilon_i$  il termine di errore.

Quando si assume che il termine di errore  $\varepsilon_i$  abbia una distribuzione normale, il modello di misura, combinato con il modello soglia, è un modello probit binario se la variabile risposta è dicotomica (Bliss, 1934) e un modello probit ordinale se la variabile risposta ha più di due modalità ordinabili (Samejima, 1969). In un modello di regressione probit le probabilità associate alle modalità categoriche si possono derivare dalla funzione di regressione  $y_i^* = \nu_i + \varepsilon_i$ , dove  $y_i^* \sim N(\nu_i, V(\varepsilon))$  e  $V(\varepsilon)$  è



standardizzato al valore uno. In questo caso le probabilità di risposta sono date dall'area sotto la curva di densità normale avente media  $\nu_i$  e varianza unitaria:

$$\Pr(y_i = k | \nu_i) = \Phi(\tau_{k+1} - \nu_i) - \Phi(\tau_k - \nu_i),$$

dove  $\Phi(\cdot)$  è la funzione di distribuzione normale standardizzata. La probabilità cumulata che una risposta assuma una modalità superiore a  $k$  è invece:

$$\Pr(y_i > k | \nu_i) = \Phi(\nu_i - \tau_{k+1}).$$

Quando si assume che il termine di errore  $\epsilon_i$  abbia una distribuzione logistica, il modello di misura è un modello logit binomiale se la variabile risposta è dicotomica e un modello logit ordinale se la variabile risposta ha più di due modalità ordinabili.

La possibilità di includere le soglie nel modello di misura, consente di mantenere la relazione lineare tra questo e le variabili latenti sottostanti,  $\mathbf{y}_i^* = \mathbf{v}_y + \mathbf{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i$  e  $\boldsymbol{\Sigma}_0^* = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ .

L'assunzione dell'esistenza di una variabile latente continua sottostante ad ogni variabile categorica permette la stima una serie di misure di correlazione latente per ciascuna coppia di variabili attraverso la funzione di ripartizione normale (Olsson 1979; Olsson, Drasgow, Dorans, 1982) oppure con il metodo di massima verosimiglianza (Lee e Poon, 1987; Poon e Lee, 1987; Bentler, 1990a,b; Joreskog 1990). In particolare, queste misure di correlazione si possono definire quando le variabili sono entrambe categoriche ordinali (correlazione policorica), entrambe dicotomiche (correlazione tetracorica), una continua e l'altra dicotomica (correlazione biseriale<sup>33</sup>), una continua e l'altra categorica (correlazione poliseriale).

Nonostante l'assunzione di normalità bivariata delle variabili risposta latenti sia stata criticata perché ritenuta non realistica (Yule, 1912; Lord e Novick, 1968), in pratica la sua convenienza è molto sostenuta (Pearson e Heron, 1913; Muthén e

---

<sup>33</sup> Quando la variabile è dicotomica per natura e non il risultato di una variabile latente sottostante, la correlazione si definisce punto biseriale.

Hofacker, 1988). Uno studio di simulazione effettuato per valutare gli effetti sul coefficiente di correlazione policorica quando l'ipotesi di normalità bivariata non è rispettata<sup>34</sup>, ha dimostrato che questo risulta appena sovrastimato<sup>35</sup> rispetto al valore del coefficiente di correlazione calcolato sulle variabili latenti non normali (Quiroga, 1992). Tuttavia, nonostante il coefficiente di correlazione policorica sia abbastanza robusto a violazioni dell'ipotesi di normalità delle variabili latenti sottostanti, non sono noti gli effetti sulla stima della matrice di covarianza asintotica tra le correlazioni policoriche né sul modello di misura né su un modello ad equazioni strutturali.

### 2.3.2 Modello di misura con variabili esplicative osservate

Una formulazione più generale del modello di misura (Muthén, 1979, 1983, 1984, 1989b) comprende anche gli effetti diretti di variabili esplicative osservate sulle variabili risposta latenti:

$$\mathbf{y}_i^* = \mathbf{v} + \underbrace{\mathbf{\Lambda}\boldsymbol{\eta}_i + \mathbf{K}\mathbf{x}_i}_{\mathbf{v}_i} + \boldsymbol{\varepsilon}_i,$$

dove  $\mathbf{v}' = (v_1, v_2, \dots, v_p)$  è un vettore  $1 \times p$  di intercette,  $\mathbf{\Lambda}$  è una matrice  $p \times m$  di coefficienti fattoriali  $(\lambda_{ij})$  che legano le variabili latenti ai rispettivi indicatori,  $\mathbf{K}$  è una matrice  $p \times q$  di coefficienti  $(k_{ij})$  che legano le variabili latenti sottostanti agli indicatori osservati alle variabili esplicative, riassunte nel vettore  $\mathbf{x}_i' = (x_1, x_2, \dots, x_q)$  mentre  $\boldsymbol{\varepsilon}' = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)$  è un vettore  $1 \times p$  di errori di misura. Di solito, solo poche righe della matrice  $\mathbf{K}$  hanno valori diversi da zero, dove un valore diverso da zero corrisponde ad una variabile  $y$  direttamente influenzata da una o più variabili esplicative. Secondo la formulazione di Muthén, il modello di misura, come il modello a variabili latenti, risulta condizionato alle variabili esplicative.

<sup>34</sup> Lo studio è stato effettuato considerando due variabili categoriche aventi ognuna quattro modalità ordinate e modificando asimmetria e curtosi delle corrispondenti variabili latenti continue.

<sup>35</sup> L'entità della distorsione è inferiore al 2%.

### 2.3.3 Le matrici implicate dei momenti con variabili risposta latenti

Il modello ad equazioni strutturali con variabili risposta osservate categoriche per le quali si ipotizza l'esistenza di una risposta latente continua,  $\mathbf{y}_i^*$ , è stato formulato da Muthén (1983, 1984). Il modello a variabili latenti,  $\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i$ , ed il modello di misura,  $\mathbf{y}_i^* = \mathbf{v}_{y^*} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{K}\mathbf{x}_i + \boldsymbol{\varepsilon}_i$ , sono condizionati al vettore di variabili esplicative,  $\mathbf{x}_i$ . Dalle equazioni che definiscono il modello ad equazioni strutturali si ricava la forma ridotta del modello di risposta:

$$\mathbf{y}_i^* = \mathbf{v}_{y^*} + \boldsymbol{\Lambda}_{y^*}(\mathbf{I} - \mathbf{B})^{-1}[\boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i] + \mathbf{K}\mathbf{x}_i + \boldsymbol{\varepsilon}_i.$$

La distribuzione delle variabili risposta latenti continue  $\mathbf{y}_i^*$  condizionata alle variabili esplicative  $\mathbf{x}_i$  si assume normale multivariata,  $\mathbf{y}_i^* \sim N_p(\cdot)$  perciò è sufficiente considerare il valore atteso e la varianza condizionati. Il vettore delle medie condizionate implicato dal modello è dato da:

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = E(\mathbf{y}_i^* | \mathbf{x}_i) = \mathbf{v}_{y^*} + \boldsymbol{\Lambda}_{y^*}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\boldsymbol{\Lambda}_{y^*}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma} + \mathbf{K})\mathbf{x}_i,$$

mentre la struttura di covarianza condizionata è data da:

$$\boldsymbol{\Sigma}_{y^*y^*}(\boldsymbol{\theta}) = V(\mathbf{y}_i^* | \mathbf{x}_i) = \boldsymbol{\Lambda}_{y^*}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Lambda}_{y^*}' + \boldsymbol{\Theta}_{\varepsilon}.$$

In presenza di variabili risposta categoriche, la scala delle variabili latenti  $\mathbf{y}^*$  è indeterminata perciò si possono considerare variabili risposta latenti standardizzate, indicate con  $\mathbf{y}_{si}^*$ . Tali variabili si ottengono moltiplicando le variabili risposta per la matrice di scala diagonale, indicata con  $\boldsymbol{\Delta}$ :

$$\mathbf{y}_{si}^* = \boldsymbol{\Delta}\mathbf{y}_i^*,$$

dove,  $\text{diag}[\mathbf{\Lambda}] = \text{diag}[V(\mathbf{y}_i^* | \mathbf{x}_i)]^{-1/2}$ .

Nell'analisi convenzionale con variabili risposta continue la matrice di scala diagonale è posta uguale alla matrice identità,  $\mathbf{\Lambda} = \mathbf{I}$ , e non ha alcun effetto. In questo caso, il vettore delle medie condizionate implicato dal modello e la struttura di covarianza condizionata assumono rispettivamente la forma seguente:

$$\begin{aligned}\boldsymbol{\mu}(\boldsymbol{\theta}) &= \mathbf{\Lambda} \mathbf{E}(\mathbf{y}_i^* | \mathbf{x}_i), \\ \boldsymbol{\Sigma}_{y^*y^*}(\boldsymbol{\theta}) &= \mathbf{\Lambda} V(\mathbf{y}_i^* | \mathbf{x}_i).\end{aligned}$$

## 2.4 La stima dei parametri

Assumendo che il modello sia identificato, è necessario stimare i parametri, riassunti per semplicità nel vettore  $\boldsymbol{\theta} = (\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Lambda}_y, \mathbf{\Lambda}_x, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Theta}_\varepsilon, \mathbf{\Theta}_\delta)$ , i cui elementi sono costituiti dalle matrici che contengono i parametri strutturali e le matrici di covarianza<sup>36</sup>.

La procedura di stima prende origine dalla relazione che formalizza l'ipotesi fondamentale dei modelli ad equazioni strutturali in base alla quale si esprime la matrice di covarianza delle variabili osservate nella popolazione,  $\mathbf{\Sigma}$ , in funzione dei parametri,  $\boldsymbol{\theta}$ . Se il modello è corretto ed i parametri sono noti, la matrice di covarianza della popolazione viene riprodotta perfettamente,  $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$ . Tuttavia, poiché non sono noti né la matrice di covarianza delle variabili osservate nella popolazione né il vettore dei parametri, la relazione fondamentale diventa  $\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$ , dove  $\hat{\mathbf{\Sigma}}$  è la matrice di covarianza implicata dal vettore dei parametri stimati, riassunti in  $\hat{\boldsymbol{\theta}}$ .

Il vettore dei parametri incogniti è stimato in modo da ottenere quei valori che rendono la matrice di covarianza campionaria  $\mathbf{S} = [s_{ij}]_{p \times p}$  più simile possibile alla matrice di covarianza implicata stimata,  $\hat{\mathbf{\Sigma}}$ . Per raggiungere questo obiettivo, si definisce una funzione di adattamento,  $F = f(\mathbf{S}, \hat{\mathbf{\Sigma}})$ , che misura la distanza tra la matrice di covarianza campionaria  $\mathbf{S}$  e la matrice di covarianza implicata dal vettore dei

<sup>36</sup> Forse, in maniera più accurata, il vettore dei parametri contenuti in  $\mathbf{\Phi}$  si potrebbe omettere, poiché contiene varianze e covarianze delle variabili esogene e non è strutturato in termini dei parametri del modello. Infatti, le stime in  $\mathbf{\Phi}$  saranno identiche ai corrispondenti elementi contenuti nella matrice di covarianza campionaria  $\mathbf{S}$ .

parametri stimati,  $\hat{\Sigma}$ . In generale, una corretta funzione di adattamento è caratterizzata dalle seguenti proprietà (Browne, 1984): i) è uno scalare, ii) assume valori reali positivi o nulli,  $F = f(\mathbf{S}, \hat{\Sigma}) \geq 0$ , e iii) assume valore nullo solo se le stime dei parametri riproducono la matrice di covarianza campionaria in maniera perfetta, cioè  $F = f(\mathbf{S}, \hat{\Sigma}) = 0$  se e solo se  $\mathbf{S} = \hat{\Sigma}$ , ed iv) è una funzione continua e differenziabile in  $\mathbf{S}$  e  $\hat{\Sigma}$ .

Il vettore dei parametri stimati  $\hat{\theta}$  si ottiene minimizzando la funzione di adattamento mediante una procedura numerica iterativa,  $\hat{\theta} = \arg \min_{\theta} f(\mathbf{S}, \hat{\Sigma})$ . Se la funzione di adattamento soddisfa le condizioni precedenti, lo stimatore è consistente (Browne, 1984). Altre proprietà importanti di una funzione di adattamento sono quelle di invarianza di scala e libertà di scala (Swaminathan e Algina, 1978) e si riferiscono alle conseguenze che derivano dalla trasformazione di una o più variabili osservate. Una funzione di adattamento è invariante rispetto alla scala se il suo valore rimane inalterato in seguito alla modifica dell'unità di misura, cioè se  $f(\mathbf{S}, \hat{\Sigma}) = f[\mathbf{DSD}, \mathbf{D}\hat{\Sigma}(\theta)\mathbf{D}]$ , dove  $\mathbf{D}$  è una matrice di scala diagonale, non singolare, avente valori positivi sulla diagonale principale<sup>37</sup>. La libertà di scala, invece, è la proprietà di mantenere un'equivalenza tra le stime dei parametri ottenute dalle variabili originali e quelle ottenute in seguito ad una trasformazione lineare. In particolare, se le variabili osservate sono sottoposte ad una trasformazione di scala, tale che  $\mathbf{y}' = \mathbf{D}_y \mathbf{y}$ , dove  $\mathbf{D}_y$  è la matrice diagonale che contiene i fattori di scala, la libertà di scala consente di ricavare la matrice dei parametri dalla matrice non trasformata, semplicemente utilizzando la matrice di scala,  $\mathbf{D}_y$ .

#### 2.4.1 Stima con variabili osservate continue distribuite normalmente

La stima dei parametri nei modelli ad equazioni strutturali nel caso di variabili risposta continue fa riferimento alla teoria normale (NT) e si basa su alcune importanti assunzioni (Bollen, 1989; Bentler e Dudgeon, 1996): i) dati generati da osservazioni

<sup>37</sup> Nel caso particolare in cui gli elementi sulla diagonale principale di  $\mathbf{D}$  siano costituiti dall'inverso degli scarti quadratici medi delle variabili osservate, allora  $\mathbf{DSD}$  diventa una matrice di correlazione. Questo significa che il valore della funzione di adattamento rimane inalterato se si utilizza la matrice di covarianza o la matrice di correlazione.

indipendenti<sup>38</sup> ed identicamente distribuite, *iid*; ii) un insieme di  $p$  variabili risposta osservate continue aventi una distribuzione normale multivariata<sup>39</sup>,  $\mathbf{y}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ; iii) un modello ipotizzato specificato in maniera corretta ed in grado di riflettere la reale struttura delle relazioni nella popolazione,  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ ; iv) un campione di grandi dimensioni; v) dati mancanti completamente a caso (MCAR); vi) esogeneità delle variabili esogene (Kaplan, 2004).

Se le condizioni sono rispettate, la matrice di covarianza campionaria  $\mathbf{S}$  contiene le statistiche sufficienti per la stima del vettore dei parametri  $\boldsymbol{\theta}$ . In questo caso le funzioni di adattamento più utilizzate sono quella di massima verosimiglianza (ML) e quella dei minimi quadrati generalizzati (GLS), i cui stimatori godono di proprietà asintotiche desiderabili.

La funzione di adattamento di massima verosimiglianza<sup>40</sup> assume la forma seguente:

$$F_{ML} = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + tr[\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] - \log|\mathbf{S}| - p,$$

dove  $tr$  indica la traccia di una matrice mentre  $p$  è il numero di variabili osservate. La matrice di covarianza asintotica dello stimatore di massima verosimiglianza  $\hat{\boldsymbol{\theta}}$  si ricava invece nel modo seguente:

---

<sup>38</sup> Tale assunzione è rispettata se le unità statistiche sono selezionate con una procedura di campionamento casuale semplice.

<sup>39</sup> Affinché le variabili osservate abbiano una distribuzione normale multivariata, le distribuzioni marginali devono avere asimmetria e curtosi di una distribuzione normale. Per valutare la forma della distribuzione multivariata, di solito si utilizzano indici di asimmetria e curtosi, univariata e multivariata (Mardia, 1970, 1974, 1985). Gli indici univariati consentono di individuare le variabili che si discostano da una distribuzione normale tuttavia definiscono solo una condizione necessaria ma non sufficiente perché la distribuzione multivariata può non essere normale anche se le variabili singolarmente considerate hanno una distribuzione normale. Per questa ragione si preferiscono gli indici multivariati.

<sup>40</sup> Il metodo della massima verosimiglianza fu proposto inizialmente in econometria per la stima dei modelli ad equazioni simultanee con il nome di “full-information maximum likelihood” (Koopmans, Rubin & Leipnik, 1950). Il punto di partenza è la funzione di verosimiglianza,  $L(\boldsymbol{\theta}; \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$  in cui il vettore dei parametri  $\boldsymbol{\theta}$  viene fatto variare mentre le variabili  $\mathbf{z}_i$  sono valori fissi perché osservati in una specifica realizzazione campionaria (Bollen, pp. 132). Definito  $\hat{\boldsymbol{\theta}}$  il suo stimatore, la stima di massima verosimiglianza è quel valore del vettore dei parametri che massimizza la probabilità (o verosimiglianza) di generare la realizzazione campionaria osservata.

$$a\text{cov}_{F_{ML}}(\hat{\boldsymbol{\theta}}) = \left\{ -E \left[ \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \right\}^{-1},$$

dove l'espressione all'interno della parentesi graffa è la matrice informazione di Fisher, indicata come  $I(\hat{\boldsymbol{\theta}})$ . L'inversa della matrice informazione è una matrice  $p \times p$  che contiene le varianze asintotiche sulla diagonale principale e le covarianza asintotiche al di fuori.

La funzione di adattamento dei minimi quadrati generalizzati<sup>41</sup> assume invece la forma seguente:

$$F_{GLS} = \frac{1}{2} \text{tr} [\mathbf{S}^{-1} (\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))^2].$$

Gli stimatori che derivano dalle funzioni di adattamento di massima verosimiglianza e dei minimi quadrati generalizzati godono delle stesse proprietà asintotiche (Browne, 1974): correttezza, consistenza, efficienza (tra gli stimatori consistenti nessuno ha una varianza asintotica più piccola) e normalità<sup>42</sup>. In particolare, quando sono rispettate le condizioni di continuità e normalità multivariata delle variabili osservate, tali funzioni sono anche invarianti alla scala e insensibili alla scala<sup>43</sup>. Inoltre, poiché è nota la matrice di covarianza asintotica delle stime, è possibile ottenere i test di significatività (Browne, 1982, 1984).

La funzione di adattamento  $F_{GLS}$  può considerarsi un caso particolare della più generale funzione di adattamento dei minimi quadrati ponderati (WLS). Tale funzione si può esprimere anche in forma quadratica e comprende le diverse funzioni di adattamento come casi particolari di una famiglia di funzioni utilizzate per l'analisi delle strutture di covarianza (Browne, 1984):

---

<sup>41</sup> Il metodo dei minimi quadrati generalizzati fu proposto da Aitken (1934) ed applicato alla path analysis da Goldberg e Jöreskog (1972). Questo metodo consente ottenere stime corrette in presenza di errori eteroschedastici.

<sup>42</sup> Tale proprietà suggerisce che per campioni numerosi, se si conosce l'errore standard delle stime dei parametri, il rapporto tra le stime ed i rispettivi errori standard approssima una distribuzione normale standard.

<sup>43</sup> La proprietà di insensibilità di scala non si mantiene se i parametri sono vincolati ad assumere valori costanti diversi da zero o se sono soggetti a restrizioni di uguaglianza o disuguaglianza (Swaminathan e Algina, 1978; Browne, 1982).

$$F_{WLS} = [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})] \mathbf{W}^{-1} [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]^{44},$$

dove  $\mathbf{s}' = (s_{11}, s_{21}, s_{22}, s_{32}, \dots, s_{pp}) = \text{vec}(\mathbf{S})$  e  $\boldsymbol{\sigma}'(\boldsymbol{\theta}) = (\sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{32}, \dots, \sigma_{pp}) = \text{vec}(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}))$  sono i vettori colonna formati dai  $\frac{1}{2}p(p+1)$  elementi non duplicati<sup>45</sup> appartenenti rispettivamente alla matrice di covarianza campionaria  $\mathbf{S}$  e alla matrice di covarianza implicata  $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$  mentre  $\mathbf{W}$  è una matrice di pesi definita positiva<sup>46</sup> di ordine  $\frac{1}{2}p(p+1) \times \frac{1}{2}p(p+1)$ .

Gli stimatori WLS si caratterizzano in relazione alla matrice di pesi utilizzata. In condizioni di regolarità molto generali (Browne, 1984), se il modello è valido nella popolazione,  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ , e la matrice di covarianza campionaria  $\mathbf{S}$  converge in probabilità ai corrispondenti elementi della matrice di covarianza della popolazione  $\boldsymbol{\Sigma}$  al crescere della numerosità campionaria, qualunque matrice di pesi  $\mathbf{W}$  definita positiva restituisce stimatori consistenti di  $\boldsymbol{\theta}$ . Tuttavia, non tutte le matrici conducono a stimatori efficienti<sup>47</sup>. Se  $\mathbf{W}$  è una stima consistente della matrice di covarianza asintotica di  $\mathbf{s}$ , allora  $F_{WLS}$  è nota come funzione dei minimi quadrati generalizzati o analisi del chi-quadrato minimo (Ferguson, 1958; Fuller, 1987; Satorra, 1992) e lo stimatore che ne deriva è asintoticamente efficiente nell'ambito delle funzioni appartenenti alla classe WLS (Browne, 1982, 1984). Nel caso particolare in cui il vettore di variabili osservate abbia una distribuzione normale multivariata o se  $\mathbf{S}$  segue una distribuzione di Wishart, la matrice di covarianza asintotica di  $\mathbf{s}$  assume una struttura particolarmente semplice e dipende solo dai momenti di secondo ordine:

$$\mathbf{W}_{NT} = 2\mathbf{K}_p' (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \mathbf{K}_p,$$

<sup>44</sup> La funzione di adattamento dei minimi quadrati ponderati si può esprimere, in maniera equivalente, anche come  $F_{WLS} = 1/2 \text{tr}[\mathbf{V}^{-1}(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta}))^2]$ , dove  $\mathbf{V}$  è una matrice di pesi diversa da  $\mathbf{W}$  ed avente una dimensione  $p \times p$  (Browne, 1974).

<sup>45</sup> Sono gli elementi della parte triangolare inferiore della matrice della matrice di covarianza campionaria, estratti per riga fino alla diagonale principale ed inseriti in un vettore colonna.

<sup>46</sup> Se la matrice  $\mathbf{W}$  non è definita positiva, la funzione non si può considerare una funzione di adattamento.

<sup>47</sup> Un esempio di questo è la funzione di adattamento  $F_{ULS}$  (minimi quadrati non ponderati), ottenuta quando  $\mathbf{W} = \mathbf{I}$ .



dove  $\mathbf{K}_p'$  e  $\mathbf{K}_p$  rappresentano matrici “eliminazione” (Magnus e Neudecker, 1988) e sono definite in Browne (1974, p. 210) mentre  $\otimes$  è il prodotto di Kronecker. Uno stimatore consistente di  $\mathbf{W}_{NT}$  si ottiene sostituendo  $\Sigma$  con  $\mathbf{S}$ . L'utilizzo di  $\mathbf{W}_{NT}$  nella funzione di adattamento  $F_{WLS}$  conduce alla funzione di adattamento  $F_{GLS}$  (Satorra, 1992).

In molte applicazioni  $\mathbf{W}$  è una matrice stocastica che converge in probabilità ad una matrice definita positiva al tendere all'infinito della numerosità campionaria o una matrice definita positiva composta da valori costanti. Quando  $\mathbf{W}$  è costituita dalla matrice identità,  $\mathbf{W} = \mathbf{I}$ , si ottiene la funzione di adattamento dei minimi quadrati non ponderati (ULS). La funzione di adattamento  $F_{ULS}$  conduce ad uno stimatore corretto e consistente di  $\boldsymbol{\theta}$  senza assumere la normalità multivariata delle variabili risposta osservate. Di conseguenza, per tale stimatore non sono fornite stime degli errori standard delle stime dei parametri né un indice di adattamento del modello<sup>48</sup>. Inoltre, tale stimatore non è il più efficiente e le stime che si ottengono sono sensibili alla scala (Schumacker e Lomax, 1996).

Un'altra possibile scelta per la matrice dei pesi  $\mathbf{W}$  si ha quando  $p \lim \mathbf{W} = \Sigma$  ed è quella in cui  $\mathbf{W} = \Sigma(\hat{\boldsymbol{\theta}})$ , dove  $\hat{\boldsymbol{\theta}}$  è il vettore dei parametri che minimizza la funzione di adattamento  $F_{ML}$ . In questo caso la matrice di pesi viene aggiornata ad ogni iterazione della procedura di stima, a differenza di quanto accade per le altre funzioni. L'utilizzo di  $\mathbf{W} = \Sigma(\hat{\boldsymbol{\theta}})$  nella funzione di adattamento  $F_{WLS}$  conduce alla funzione di adattamento  $F_{ML}$ <sup>49</sup>.

#### 2.4.2 Effetti con variabili osservate non distribuite normalmente

Il comportamento degli stimatori basati sulla teoria normale (ML e GLS) in presenza di variabili osservate aventi una distribuzione multivariata non normale è stato oggetto di diversi studi di simulazione ed i risultati sono ormai piuttosto noti in letteratura. In

<sup>48</sup> Per tale funzione di adattamento sono stati comunque suggeriti metodi per il calcolo dei test di significatività delle stime (Browne, 1982).

<sup>49</sup> Dal punto di vista tecnico, in questo modo si ottiene la funzione di adattamento RLS (Reweighted Least Squares) che è asintoticamente equivalente alla funzione di adattamento  $F_{ML}$ .

generale, gli effetti sulle stime aumentano al crescere del grado di non normalità.

Le stime dei parametri rimangono consistenti anche in presenza di variabili aventi distribuzioni non normali (Bollen, 1989; Finch, West e MacKinnon, 1997), indipendentemente dallo stimatore utilizzato,  $F_{ML}$ ,  $F_{GLS}$  e  $F_{ULS}$  (Bollen, 1989). In presenza di variabili con distribuzioni ellittiche multivariate<sup>50</sup>, gli stimatori appartenenti alla classe GLS mantengono anche l'efficienza asintotica (Browne, 1984).

Gli errori standard delle stime dei parametri risultano invece sottostimati (Chou, Bentler e Satorra, 1991; Finch, West e MacKinnon, 1997; Olsson, Foss, Troye e Howell, 2000) e questo comporta errori di I tipo più frequenti per i test di significatività<sup>51</sup>. In particolare, l'effetto sugli errori standard è maggiore in presenza di distribuzioni leptocurtiche<sup>52</sup> (Hoogland e Boomsma, 1998).

In definitiva, in assenza di normalità multivariata delle variabili osservate, si mantiene solo la consistenza degli stimatori basati sulle funzioni di adattamento  $F_{ML}$  e  $F_{GLS}$  ma la curtosi eccessiva elimina l'efficienza asintotica e rende la matrice di covarianza asintotica stimata potenzialmente inesatta, influenzando i tradizionali test di significatività.

#### 2.4.3 *Stima con variabili continue non distribuite normalmente*

L'assunzione di continuità e normalità multivariata delle variabili osservate è difficilmente soddisfatta e gli stimatori basati sulla teoria normale (NT) non sono robusti in presenza di variabili con distribuzione non normale.

La teoria classica di analisi delle strutture di covarianza è stata generalizzata da Browne (1982, 1984) ad una distribuzione multivariata per variabili continue che soddisfa assunzioni molto deboli e si basa sulla funzione di adattamento dei minimi quadrati ponderati (WLS). In questo caso, la matrice di covarianza asintotica di  $\mathbf{s}$ , indicata con  $\mathbf{\Gamma}$ , è composta dai seguenti elementi (Browne, 1982):

---

<sup>50</sup> Le variabili aventi distribuzioni ellittiche multivariate sono prive di asimmetria ma presentano lo stesso grado di curtosi.

<sup>51</sup> Questo comporta che siano considerati significativamente diversi da zero e quindi inclusi nel modello parametri con valore nullo nella popolazione.

<sup>52</sup> Sono distribuzioni che presentano curtosi positiva, caratterizzate da una forma più appuntita rispetto alla distribuzione normale.

$$\gamma_{ijkl} = \sigma_{ijkl} - \sigma_{ij}\sigma_{kl},$$

dove  $\sigma_{ijkl} = E(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l)$  è il momento centrale di quarto ordine mentre  $\sigma_{ij}$  e  $\sigma_{kl}$  sono le matrici di covarianza nella popolazione rispettivamente di  $x_i$  con  $x_j$  e di  $x_k$  con  $x_l$ . Si definisce il vettore di dati  $\mathbf{d}_i$  riferito all'osservazione  $i$  ed avente dimensioni  $\frac{1}{2}p(p+1)$ :

$$\mathbf{d}_i = \begin{pmatrix} (y_{i1} - \bar{y}_1)(y_{i1} - \bar{y}_1) \\ (y_{i2} - \bar{y}_2)(y_{i1} - \bar{y}_1) \\ (y_{i2} - \bar{y}_2)(y_{i2} - \bar{y}_2) \\ \dots \\ (y_{ip} - \bar{y}_p)(y_{ip} - \bar{y}_p) \end{pmatrix},$$

dove  $y_{iv}$  è la  $i$ -esima osservazione rispetto alla variabile  $v$  ( $v=1,2,\dots,p$ ) e  $\bar{y}_v$  è la corrispondente media. Per un campione di  $n$  unità, il vettore  $\mathbf{s}$  si ottiene nel modo seguente:

$$(n-1)^{-1} \sum_{i=1}^n \mathbf{d}_i = \mathbf{s}.$$

Uno stimatore consistente di  $\mathbf{\Gamma}$  si può ottenere mediante la matrice di covarianza campionaria di  $\mathbf{d}_i$  che coinvolge i momenti di quarto ordine (Browne, 1982; 1984; Chamberlain, 1982; Satorra, 1992):

$$\hat{\mathbf{\Gamma}} \equiv (n-1)^{-1} \sum_{i=1}^n (\mathbf{d}_i - \bar{\mathbf{d}})(\mathbf{d}_i - \bar{\mathbf{d}})',$$

in modo tale che lo stimatore della matrice di covarianza asintotica di  $\mathbf{s}$  sia  $n^{-1}\hat{\mathbf{\Gamma}}$ .

Considerando  $\hat{\mathbf{\Gamma}} = \mathbf{W}$  nella funzione di adattamento dei minimi quadrati ponderati, si ottiene lo stimatore definito “*asymptotically distribution free best GLS estimators*” (Browne, 1984) o più semplicemente “*arbitrary distribution free*” (ADF)

per la sua proprietà di valere in condizioni generali, senza cioè specificare alcuna forma della distribuzione delle variabili osservate. In questo caso, la matrice di covarianza asintotica delle stime dei parametri si può scrivere nel modo seguente:

$$a \text{ cov}(\hat{\boldsymbol{\theta}}) = n^{-1} (\boldsymbol{\Delta}' \mathbf{W}^{-1} \boldsymbol{\Delta})^{-1},$$

dove  $\boldsymbol{\Delta} = \partial \boldsymbol{\sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$  è la matrice di derivate di ordine  $1/2 p(p+1)q$ . Tale espressione viene utilizzata comunemente per gli stimatori NTGLS e ADF ma Satorra (1992) ha dimostrato che la stessa espressione vale asintoticamente anche per lo stimatore ML.

Quando le variabili osservate hanno una distribuzione normale multivariata ed il modello è specificato in maniera corretta,  $\sigma_{ijkl} = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}$  e gli elementi della matrice di covarianza asintotica di  $\mathbf{s}$  delle matrici  $\mathbf{W}_{WLS}$  e  $\mathbf{W}_{GLS}$  convergono in probabilità ai valori della popolazione,  $\gamma_{ijkl} = \sigma_{ig}\sigma_{jh} + \sigma_{ih}\sigma_{jg}$ , sebbene la loro variabilità sia diversa (Hu, Bentler e Kano, 1992). In questo caso gli stimatori ML, GLS e WLS convergono allo stesso vettore ottimo dei parametri e presentano le stesse proprietà asintotiche (Browne, 1974, 1984).

Nonostante lo stimatore ADF sia attraente dal punto di vista teorico, in pratica presenta alcuni problemi computazionali che ne limitano le possibilità applicative. Il principale riguarda l'intensità di calcolo necessaria ottenere l'inversa della matrice dei pesi  $\mathbf{W}$ . In particolare, con  $p$  variabili osservate, la matrice di covarianza è composta da  $u = \frac{1}{2} p \times (p+1)$  elementi mentre la matrice  $\mathbf{W}$  è di ordine  $u \times u$  e cresce rapidamente all'aumentare del numero di variabili osservate. Inoltre, quando il numero di variabili è molto elevato, spesso la matrice dei pesi non è definita positiva e quindi non può essere invertita (Bentler, 1995; West, Finch e Curran, 1995). In questo caso, per assicurare una matrice  $\mathbf{W}$  non singolare, è necessaria una numerosità campionaria di almeno  $p + \frac{1}{2} p(p+1)$  unità.

Oltre ai problemi di tipo computazionale, le ricerche empiriche sulle prestazioni dello stimatore ADF mostrano risultati non proprio soddisfacenti, soprattutto per modelli di moderate o grandi dimensioni e/o campioni di numerosità limitata. Le stime dei parametri tendono ad assumere una distorsione negativa, a meno che il campione osservato non sia sufficientemente grande. Il livello di distorsione tende ad essere più

pronunciato al crescere del grado di curtosi. Le stime degli errori standard di un modello correttamente specificato sono invece maggiori rispetto alle stime ML solo quando le variabili osservate presentano una curtosi univariata media maggiore di tre e la numerosità campionaria è maggiore di 400 unità (Hoogland e Boomsma, 1998).

Per quanto riguarda gli indici di adattamento del modello, lo stimatore ADF fornisce risultati errati a meno che la numerosità campionaria non sia estremamente grande (Olsson, Foss, Troye e Howell, 2000). Tale stimatore produce tassi di errore errori di I tipo accettabili solo quando la numerosità è di almeno 5000 unità (Hu, Bentler e Kano, 1992). Analogamente, le misure basate sul  $\chi^2$  tendono ad aumentare al diminuire della numerosità campionaria e/o del grado di non normalità, con la conseguenza di rifiutare troppo frequentemente modelli specificati correttamente (Curran, West e Finch, 1996). Anche più problematica appare la mancanza di sensibilità dello stimatore ADF in presenza di modelli mal specificati con la conseguenza che in tali situazioni diventa meno probabile riuscire a respingere modelli mal specificati. Tale problema risulta aggravato all'aumentare del livello di scostamento (Curran, West e Finch, 1996; Olsson, Foss, Troye e Howell, 2000).

Ispirandosi al lavoro di Satorra (1992), Muthén (1993) ha proposto un approccio robusto alternativo per il calcolo della matrice di covarianza asintotica delle stime dei parametri e dei test di adattamento del modello. Tale approccio prevede la stima del vettore di parametri  $\theta$  con la funzione di adattamento dei minimi quadrati ponderati<sup>53</sup>, in cui l'espansione in serie di Taylor restituisce la matrice di covarianza asintotica seguente (Ferguson, 1958; Browne, 1984, Fuller, 1987, Satorra, 1992):

$$a\text{cov}(\hat{\theta}) = n^{-1}(\Delta'W^{-1}\Delta)^{-1}\Delta'W^{-1}\Gamma W^{-1}\Delta(\Delta'W^{-1}\Delta)^{-1}.$$

La matrice di covarianza asintotica fornisce una stima robusta degli errori standard dei parametri. Una stima consistente di questa matrice si può ottenere considerando  $\Delta = \Delta(\hat{\theta})$  e sostituendo  $\Gamma$  con la sua stima consistente,  $\hat{\Gamma}$ .

L'espressione generale della matrice di covarianza asintotica delle stime,  $a\text{cov}(\hat{\theta}) = n^{-1}(\Delta'W^{-1}\Delta)^{-1}\Delta'W^{-1}\Gamma W^{-1}\Delta(\Delta'W^{-1}\Delta)^{-1}$ , dimostra che non è necessario avere

---

<sup>53</sup> Per i riferimenti generali alla teoria di base, si consultino Browne (1982; 1984) e Satorra (1989; 1992).

una matrice di pesi  $\mathbf{W} = \hat{\mathbf{\Gamma}}$  ma che si può anche scegliere  $\mathbf{W}_{NT} = 2\mathbf{K}_p'(\mathbf{\Sigma} \otimes \mathbf{\Sigma})\mathbf{K}_p$  per ottenere le stime dei parametri nel caso normale. Quando si utilizza  $\hat{\mathbf{\Gamma}} \equiv (n-1)^{-1} \sum_{i=1}^n (\mathbf{d}_i - \bar{\mathbf{d}})(\mathbf{d}_i - \bar{\mathbf{d}})'$  nell'espressione generale della matrice di covarianza asintotica, si ottiene una matrice di covarianza delle stime adeguata anche in assenza di normalità delle variabili osservate. In presenza di variabili osservate aventi una distribuzione non normale, le stime ottenute in questo modo hanno una variabilità asintotica maggiore rispetto a quella che si ottiene con lo stimatore ADF. Tuttavia, tale approccio rimane preferibile rispetto allo stimatore ADF da un punto di vista computazionale quando le variabili osservate sono piuttosto numerose. In questo modo, a differenza dello stimatore ADF, la matrice  $\mathbf{\Gamma}$  non deve essere invertita. Inoltre, la matrice  $\mathbf{W}$  si può scegliere una matrice semplice da invertire<sup>54</sup>.

Più recentemente, Asparouhov e Muthén (2003) hanno proposto tre stimatori di massima verosimiglianza (ML) basati sull'algoritmo EM. Tali stimatori si differenziano per il metodo di calcolo degli errori standard e forniscono un test di adattamento robusto basato sulla statistica chi-quadrato ed errori standard corretti in presenza di variabili osservate aventi distribuzioni non normali. Tale approccio si basa sulla funzione di log-verosimiglianza individuale proposta da Arbuckle (1996) per la stima di massima verosimiglianza completa in presenza di dati mancanti. Seguendo l'approccio di Arbuckle, la funzione di log-verosimiglianza si ottiene come somma delle funzioni di log-verosimiglianza riferite a ciascuna unità osservata, indicata con  $i$ :

$$\log L = \sum_{i=1}^n \log L_i = \sum_{i=1}^n K_i - \frac{1}{2} \log |\mathbf{\Sigma}_i| - \frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_i)' \mathbf{\Sigma}_i^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_i),$$

dove  $L_i$  è la funzione di verosimiglianza riferita all'unità  $i$ ,  $K_i$  è una costante che dipende dal numero di valori osservati rispetto all'unità  $i$ ,  $\mathbf{x}_i$  il vettore di variabili osservate a livello individuale, mentre  $\boldsymbol{\mu}_i$  e  $\mathbf{\Sigma}_i$  contengono rispettivamente le medie e la matrice di covarianza delle variabili osservate nella popolazione.

Il primo stimatore, definito con l'acronimo MLF, calcola la matrice di

---

<sup>54</sup> Muthén (1993) ha utilizzato  $\mathbf{W} = \mathbf{I}$ .

covarianza asintotica delle stime utilizzando una approssimazione del primo ordine della matrice informazione di Fisher, ottenuta nel modo seguente:

$$I_{MLF} = \sum_{i=1}^n \frac{\partial \log L_i}{\partial \boldsymbol{\theta}} \times \frac{\partial \log L_i}{\partial \boldsymbol{\theta}'}.$$

Il secondo è invece uno stimatore ML convenzionale che approssima la matrice informazione di Fisher utilizzando le derivate del secondo ordine della funzione osservata di log-verosimiglianza:

$$I_{ML} = - \sum_{i=1}^n \frac{\partial^2 \log L_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}.$$

Il terzo stimatore, definito con l'acronimo MLR, approssima la matrice informazione di Fisher utilizzando uno stimatore sandwich derivato dalle matrici informazione degli stimatori ML e MLF:

$$I_{MLR} = I_{ML}^{-1} \times I_{MLF}^{-1} \times I_{ML}^{-1}.$$

Lo stimatore MLR produce la matrice di covarianza delle stime corretta che non dipende dall'assunzione di normalità ed un test di adattamento robusto basato sulla statistica chi-quadrato.

#### 2.4.4 Effetti con variabili osservate categoriche

Le variabili osservate nell'ambito delle scienze sociali sono spesso di natura categorica, con modalità ordinate o non ordinate. Le variabili categoriche non si possono, a rigore, considerare distribuite normalmente poiché le modalità non sono continue ma discrete (Muthén e Kaplan, 1985).

Quando si analizzano variabili che presentano modalità ordinate, spesso la natura categorica delle variabili viene ignorata ed i dati sono analizzati considerando le modalità come continue. In questo caso, i metodi di stima che assumono un modello di misura lineare (ML, GLS o ULS) sono basati sulla matrice di covarianza campionaria

tra le variabili osservate categoriche e producono risultati generalmente non attendibili, soprattutto se il numero di modalità è limitato (inferiore a cinque). In generale, il livello di approssimazione che si ottiene considerando continue variabili categoriche migliora all'aumentare del numero di modalità assunte dalle variabili categoriche (Bollen, 1989) poiché i coefficienti di correlazione stimati tendono ad assumere valori più vicini a quelli reali. Invece, minore è il numero di modalità, più forte è il livello di attenuazione dei coefficienti di correlazione e quindi il divario tra il valore stimato e quello reale. Gli effetti sulle stime dei parametri, dei relativi errori standard e degli indici di adattamento sono diversi a seconda della forma della distribuzione delle variabili osservate.

Quando le variabili categoriche hanno una forma approssimativamente normale, i coefficienti fattoriali ottenuti con il metodo di massima verosimiglianza (ML) sono appena sottostimati (Babakus, Ferguson e Jöreskog, 1987; Muthén e Kaplan, 1985) mentre le stime degli errori standard sono più sensibili e mostrano una distorsione negativa più accentuata (Muthén e Kaplan, 1985; 1992; Babakus, Ferguson e Jöreskog, 1987; West, Finch e Curran, 1995) che produce test di significatività dei parametri più ampi e quindi errori di primo tipo più frequenti. Il livello di sottostima dei parametri e degli errori standard cresce al diminuire del numero di modalità, anche quando le variabili hanno una distribuzione (forma) simmetrica. Quando la distribuzione delle variabili categoriche non è approssimativamente normale, la distorsione negativa delle stime ML dei parametri diviene più pronunciata all'aumentare del livello di asimmetria e curtosi univariata mentre le stime degli errori standard si ampliano (Babakus, Ferguson e Jöreskog, 1987; Muthén e Kaplan, 1985). Il livello di distorsione aumenta al diminuire della dimensione campionaria, del numero di modalità, del grado di non normalità e quando la relazione tra i fattori e gli indicatori è più debole (Babakus, Ferguson e Jöreskog, 1987; Bollen, 1989; Dolan, 1994).

#### *2.4.5 Stima con variabili osservate categoriche*

Quando le variabili risposta osservate sono di natura categorica, sostituire la matrice di correlazione policorica con la matrice di covarianza nella funzione di adattamento di massima verosimiglianza  $F_{ML}$  è una strategia inadeguata. Infatti, anche se tale approccio restituisce di solito stime consistenti dei parametri, i test statistici e gli errori



standard non sono corretti (Babakus, Ferguson e Jöreskog, 1987; Rigdon e Ferguson, 1991).

I metodi di stima in assenza di normalità multivariata prevedono l'utilizzo di una funzione di adattamento appartenente alla classe dei minimi quadrati ponderati (WLS) sia nel caso di variabili continue (Browne, 1982, 1984) che nel caso di variabili categoriche (Muthén, 1978, 1983, 1984; Lee, Poon e Bentler, 1990, 1992; Jöreskog, 1994; Muthén e Satorra, 1995). Tali metodi utilizzano la matrice di correlazione policorica insieme alla funzione di adattamento dei minimi quadrati ponderati (generalizzati)  $F_{WLS}$  e differiscono per il modo in cui sono stimati i parametri (soglie, coefficienti di correlazione tra le variabili risposta latenti e coefficienti del modello) e per il numero di stadi richiesti dalla procedura di stima.

Gli approcci di Muthén (1983; 1984; 1995) e Jöreskog (1994) si basano su una procedura a tre stadi<sup>55</sup> in cui nella prima fase si stimano le soglie dalle distribuzioni normali univariate, nella seconda si stimano i coefficienti di correlazione latente dalla distribuzione normale bivariata, dati i valori delle soglie, e nella terza si stimano i parametri del modello con una funzione di adattamento basata sul metodo dei minimi quadrati generalizzati in cui la matrice dei pesi è costituita da una stima della matrice di covarianza asintotica dei coefficienti di correlazione latente<sup>56</sup>.

Il metodo di Lee, Poon e Bentler (1990, 1992) è basato su una procedura a due stadi che prevede nella prima fase la stima simultanea delle soglie e dei coefficienti di correlazione latente con il metodo di massima verosimiglianza e nella seconda fase la stima dei parametri del modello con il metodo dei minimi quadrati generalizzati, in base alla distribuzione congiunta asintotica degli stimatori ottenuti nella prima fase e ad un'adeguata matrice di pesi.

Nel paragrafo successivo viene descritto l'approccio ad informazione limitata proposto da Muthén, i cui risultati saranno utili anche per comprendere la stima dei modelli ad equazioni strutturali multilivello.

---

<sup>55</sup> I lavori di Muthén e Jöreskog sono stati realizzati in maniera parallela ma indipendente.

<sup>56</sup> Sebbene tali metodi siano molti simili nella stima della matrice di covarianza asintotica, differiscono per il trattamento dei parametri soglia. Uno studio di simulazione realizzato da Dolan (1994) ha dimostrato che, nel caso di un modello fattoriale confermativo, i risultati ottenuti con i due approcci sono virtualmente identici (Flora e Curran, 2004).

#### 2.4.6 Stima ad informazione limitata

L'approccio di Muthén (1984) consente la stima dei modelli ad equazioni strutturali con variabili latenti continue e variabili risposta continue, categoriche ordinali, dicotomiche, censurate e combinazioni tra queste. La metodologia proposta da Muthén estende alle variabili risposta categoriche ordinali il lavoro di Muthén e Christoffersson (1978) proposto per l'analisi fattoriale con variabili risposta binarie e generalizza il tradizionale modello LISREL, inizialmente limitato ai modelli ad equazioni strutturali con variabili risposta continue.

La stima si basa su una procedura ad informazione limitata a tre stadi per ottenere le stime dei parametri, degli errori standard e di una misura di adattamento del modello basata sulla statistica del chi quadrato. I primi due stadi sono necessari per stimare i parametri soglia, medie e varianze, i coefficienti di correlazione tra coppie di variabili risposta latenti e la matrice di covarianza asintotica tra questi. Il terzo stadio prevede invece la stima dei parametri del modello utilizzando una funzione di adattamento basata sul metodo dei minimi quadrati ponderati, avente come matrice dei pesi l'inversa della matrice di covarianza asintotica dei parametri stimati nei primi due stadi.

Il modello generale si basa sulle seguenti equazioni riferite rispettivamente al modello a variabili latenti,  $\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i$ , ed al modello di misura,  $\mathbf{y}_i^* = \mathbf{v}_{y^*} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{K}\mathbf{x}_i + \boldsymbol{\varepsilon}_i$ , entrambi condizionati al vettore di variabili esplicative,  $\mathbf{x}_i$ . Dalle equazioni che definiscono il modello ad equazioni strutturali generale si ricava la forma ridotta del modello di risposta:

$$\mathbf{y}_i^* = \mathbf{v}_{y^*} + \boldsymbol{\Lambda}_{y^*}(\mathbf{I} - \mathbf{B})^{-1}[\boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i] + \mathbf{K}\mathbf{x}_i + \boldsymbol{\varepsilon}_i.$$

Il vettore delle medie condizionate implicato dal modello è invece dato da:

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = E(\mathbf{y}_i^* | \mathbf{x}_i) = \mathbf{v}_{y^*} + \boldsymbol{\Lambda}_{y^*}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\boldsymbol{\Lambda}_{y^*}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma} + \mathbf{K})\mathbf{x}_i,$$

mentre la struttura di covarianza condizionata è data da:

$$\Sigma_{y^*y^*}(\theta) = V(y_i^* | x_i) = \Lambda_{y^*}(\mathbf{I} - \mathbf{B})^{-1} \Psi(\mathbf{I} - \mathbf{B})'^{-1} \Lambda_{y^*}' + \Theta_\varepsilon.$$

La struttura parametrica del modello generale si può suddividere, per comodità, in tre parti riassunte rispettivamente nei tre vettori  $\sigma_1$ ,  $\sigma_2$  e  $\sigma_3$ .

La prima parte contiene la struttura parametrica per medie, soglie o intercette ed è sintetizzata nel vettore  $\sigma_1$ :

$$\sigma_1 = \Lambda^* \left( \mathbf{K}_\tau \tau - \mathbf{K}_v \left( \mathbf{v} + \Lambda(\mathbf{I} - \mathbf{B})^{-1} \alpha \right) \right),$$

dove  $\Lambda^*$  è una matrice che contiene gli stessi elementi della matrice di scala diagonale  $\Lambda$  ma con componenti duplicati per le variabili categoriche che hanno più di un parametro soglia (variabili con più di due modalità),  $\mathbf{K}_\tau$  è una matrice che seleziona gli elementi dal vettore  $\tau$  per le variabili categoriche ma non per le variabili continue<sup>57</sup>,  $\tau$  contiene i parametri soglia per tutte le variabili categoriche e  $\mathbf{K}_v$  è una matrice di selezione che sceglie le righe dal vettore che pre moltiplica.

La seconda parte contiene la struttura parametrica dei coefficienti e viene utilizzata se almeno una variabile è categorica:

$$\sigma_2 = \text{vec}(\Lambda \Lambda (\mathbf{I} - \mathbf{B})^{-1} \Gamma),$$

dove l'operatore  $\text{vec}(\cdot)$  dispone gli elementi di una matrice in un vettore di dimensioni  $(pq \times 1)$ , considerando solo gli elementi della diagonale inferiore di una matrice simmetrica.

La terza parte contiene una struttura parametrica di covarianza, correlazione o correlazione residua. In presenza di variabili categoriche,

$$\sigma_3 = \mathbf{K}_\sigma \text{vec}(\Lambda (\Lambda (\mathbf{I} - \mathbf{B})^{-1} \Psi (\mathbf{I} - \mathbf{B})'^{-1} \Lambda' + \Theta) \Lambda),$$

---

<sup>57</sup> In pratica, tale matrice ha una riga di valori nulli per ciascuna variabile continua.

dove  $\mathbf{K}_\sigma$  seleziona gli elementi dal vettore che pre moltiplica in modo tale che gli elementi non sono inclusi se la variabile  $y$  corrispondente è categorica.

In generale, ciascuna variabile  $y_i$  contribuisce agli elementi di  $\sigma_1$  con  $C_i - 1$  parametri soglia se è categorica e con un solo parametro, la media, se è continua. Ciascuna variabile  $y_i$  contribuisce invece agli elementi di  $\sigma_2$  con  $q$  elementi. La coppia di variabili  $y_i$  e  $y_j$  contribuisce agli elementi di  $\sigma_3$  con un unico elemento se entrambe le variabili sono categoriche (un coefficiente di correlazione) e può contribuire con un elemento in più per ciascuna variabile  $y$  continua (una varianza). Qualunque modello stimato in base a questa struttura generale è identificato se e solo se i suoi parametri sono identificati in termini dei vettori  $\sigma_1$ ,  $\sigma_2$  e  $\sigma_3$ .

Il primo stadio della procedura prevede la stima consistente di massima verosimiglianza degli elementi di  $\sigma_1$  (parametri soglia o intercette) e  $\sigma_2$  (coefficienti di regressione), indicate rispettivamente con  $\mathbf{s}_1$  e  $\mathbf{s}_2$ . In pratica, se  $F_i(i=1,2,\dots,p)$  è la funzione di log-verosimiglianza riferita alla variabile  $y_i$  condizionata alle variabili esplicative  $x_i$ , la massimizzazione separata di ogni funzione  $F_i$  restituisce le stime di massima verosimiglianza ad informazione limitata degli elementi di  $\sigma_{1i}$  e  $\sigma_{2i}$  mediante  $p$  regressioni probit ordinate multinomiali univariate di  $y$  rispetto ad  $x$ . Tali stime sono indicate rispettivamente con  $\mathbf{s}_{1i}$  e  $\mathbf{s}_{2i}$  ed assemblate nei vettori  $\mathbf{s}_1$  e  $\mathbf{s}_2$ .

Il secondo stadio della procedura prevede la stima di pseudo massima verosimiglianza degli elementi di  $\sigma_3$ , indicata con  $\mathbf{s}_3$ , date le stime di  $\sigma_1$  e  $\sigma_2$  fissate ai valori  $\mathbf{s}_1$  e  $\mathbf{s}_2$  ottenuti al primo stadio della procedura. In pratica, se  $F_{ij}(i=1,2,\dots,p; j=1,2,\dots,p; j < i)$  è la funzione di log-verosimiglianza bivariata di  $y_i$  e  $y_j$  condizionata alle variabili esplicative  $x_i$  e alle stime ottenute massimizzando le funzioni  $F_i$ , la massimizzazione separata di ogni funzione  $F_{ij}$  restituisce le stime condizionate di pseudo massima verosimiglianza ad informazione limitata degli elementi di  $\sigma_{3ij}$  mediante regressioni probit ordinate multinomiali bivariate condizionate per ciascuna combinazione di variabili  $y_i, y_j, i \neq j$ . Tali stime sono indicate con  $\mathbf{s}_{3ij}$  ed assemblate nel vettore  $\mathbf{s}_3$ .

Il vettore delle derivate prime corrispondenti ai primi due stadi della procedura è dato da:

$$\frac{\partial F}{\partial \sigma'} = \left( \frac{\partial F_1}{\partial \sigma_{11}}, \frac{\partial F_1}{\partial \sigma_{21}}, \frac{\partial F_1}{\partial \sigma_{12}}, \frac{\partial F_1}{\partial \sigma_{22}}, \dots, \frac{\partial F_p}{\partial \sigma_{1p}}, \frac{\partial F_p}{\partial \sigma_{2p}}, \frac{\partial F_{21}}{\partial \sigma_{321}}, \frac{\partial F_{31}}{\partial \sigma_{331}}, \dots, \frac{\partial F_{pp-1}}{\partial \sigma_{3pp-1}} \right).$$

I primi due stadi dello stimatore proposto si definiscono imponendo  $\partial F / \partial \sigma = 0$ .

Muthén e Satorra (1995) hanno dimostrato che sotto le solite condizioni di regolarità<sup>58</sup>, il vettore delle stime  $\hat{\sigma}$  ricavate dai primi due stadi della procedura di stima è una stima consistente di  $\sigma$ , cioè  $p \lim \hat{\sigma} = \bar{\sigma}$ , dove  $\bar{\sigma}$  indica il “valore vero” di  $\sigma$ .

Muthén (1984) ricava lo stimatore consistente della matrice di covarianza asintotica delle stime  $\hat{\sigma}$  in base alla seguente approssimazione per grandi campioni:

$$\frac{\partial F}{\partial \sigma} + \frac{\partial^2 F}{\partial \sigma \partial \sigma'} (\hat{\sigma} - \sigma) \approx 0,$$

e, riscritta  $\partial F / \partial \sigma$  come segue:

$$\frac{\partial F}{\partial \sigma} = \sum_{r=1}^N \frac{\partial F(r)}{\partial \sigma},$$

dove  $r$  scorre rispetto alle unità osservate in un campione di numerosità  $N$ . Con il partizionamento:

$$\frac{\partial^2 F}{\partial \sigma \partial \sigma'} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix},$$

$\mathbf{A}_{11}$  è una matrice diagonale a blocchi che coinvolge le funzioni di log-verosimiglianza univariate,  $F_i$ , e  $\mathbf{A}_{22}$  è una matrice diagonale le cui righe corrispondono alle funzioni di

---

<sup>58</sup> Le condizioni possono non rappresentare l'insieme più stringente di assunzioni sotto le quali i risultati sono verificati, tuttavia sono un insieme di condizioni sufficienti.

log-verosimiglianza bivariata,  $F_{ij}$ , mentre  $\mathbf{A}_{12} = 0$ . Ogni matrice sulla diagonale di  $\mathbf{A}_{11}$  può essere approssimata nel modo seguente:

$$\sum_{r=1}^n \begin{bmatrix} \partial^2 F_i(r) / \partial \sigma_{1i} \\ \partial^2 F_i(r) / \partial \sigma_{2i} \end{bmatrix} \begin{bmatrix} \partial^2 F_i(r) / \partial \sigma_{1i}' & \partial^2 F_i(r) / \partial \sigma_{2i}' \end{bmatrix},$$

e gli elementi non nulli della matrice  $\mathbf{A}_{21}$  possono essere approssimati come ( $s = i$  o  $j$ ):

$$\sum_{r=1}^n \partial^2 F_{ij}(r) / \partial \sigma_{3ij} \begin{bmatrix} \frac{\partial^2 F_{ij}(r)}{\partial \sigma_{1s}'} & \frac{\partial^2 F_{ij}(r)}{\partial \sigma_{2s}'} \end{bmatrix},$$

e gli elementi diagonali di  $\mathbf{A}_{22}$  come:

$$\sum_{r=1}^n \left( \partial^2 F_{ij}(r) / \partial \sigma_{3ij} \right)^2.$$

Se  $\mathbf{B}$  rappresenta l'approssimazione delle matrici  $\mathbf{A}$ , segue che uno stimatore consistente della matrice di covarianza asintotica di  $\hat{\boldsymbol{\theta}}$  è dato da:

$$a \text{cov}(\hat{\boldsymbol{\theta}}) = \hat{\mathbf{B}}^{-1} \sum_{r=1}^n \frac{\partial F}{\partial \boldsymbol{\sigma}} \frac{\partial F}{\partial \boldsymbol{\sigma}'} \hat{\mathbf{B}}^{-1'},$$

calcolato a  $\boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}}$ .

Il terzo stadio della procedura prevede la stima dei parametri del modello  $\boldsymbol{\theta}$  minimizzando la funzione di adattamento dei minimi quadrati ponderati:

$$F_{WLS} = [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})]' \mathbf{W}^{-1} [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})],$$

dove  $\mathbf{W}^{-1}$  è la matrice definita positiva dei pesi. Quando la matrice dei pesi assume la forma della matrice di covarianza asintotica delle stime  $\mathbf{W} = \boldsymbol{\Gamma}$ , si ottiene uno stimatore dei parametri ad informazione limitata, la cui distribuzione è asintoticamente normale.

L'ottimizzazione della funzione di adattamento viene effettuata con il metodo quasi-Newton che richiede solo il calcolo di derivate prime e la costruzione di una matrice delle derivate seconde approssimata. Lo stimatore è consistente se il modello è specificato correttamente.

Muthén (1984) fornisce anche lo stimatore consistente della matrice di covarianza asintotica delle stime:

$$a\text{cov}(\hat{\boldsymbol{\theta}}) = n^{-1}(\boldsymbol{\Delta}'\mathbf{W}^{-1}\boldsymbol{\Delta})^{-1},$$

dove  $\boldsymbol{\Delta} = \partial\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ . Muthén e Satorra (1995) hanno dimostrato la normalità asintotica della matrice di covarianza delle stime dei parametri. In generale, per i dettagli relativi ai primi due stadi della procedura di stima si può fare riferimento ad Olsson (1979) per il calcolo dei coefficienti di correlazione policorica, ad Olsson, Dragsow e Dorans (1982) per il calcolo del coefficiente di correlazione poliseriale e a Muthén (1989c) per il coefficiente di correlazione tobit.

Muthén, du Toit e Spisic (1997) hanno esteso al modello ad equazioni strutturali proposto da Muthén (1984) l'approccio robusto per il calcolo della matrice di covarianza asintotica delle stime dei parametri e dei test di adattamento del modello utilizzato da Satorra (1990, 1992) per variabili risposta continue e da Muthén (1993) per variabili risposta binarie. Come già descritto (paragrafo 2.4.2), tale approccio prevede la stima del vettore di parametri  $\boldsymbol{\theta}$  con la funzione di adattamento dei minimi quadrati ponderati, in cui l'espansione in serie di Taylor restituisce la matrice di covarianza asintotica, che fornisce una stima robusta degli errori standard dei parametri:

$$a\text{cov}(\hat{\boldsymbol{\theta}}) = n^{-1}(\boldsymbol{\Delta}'\mathbf{W}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W}^{-1}\boldsymbol{\Gamma}\mathbf{W}^{-1}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{W}^{-1}\boldsymbol{\Delta})^{-1},$$

dove  $\boldsymbol{\Delta} = \partial\boldsymbol{\sigma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$  è la matrice di derivate di ordine  $1/2 p(p+1)q$ . Tale stimatore si caratterizza anche per la scelta della matrice dei pesi  $\mathbf{W}$ . A fini computazionali, è preferibile una matrice piuttosto semplice e facilmente invertibile. Muthén (1993) proponeva  $\mathbf{W} = \mathbf{I}$  che riduce la funzione di adattamento dei minimi quadrati ponderati ad una funzione di adattamento dei minimi quadrati non ponderati. Muthén, du Toit e

Spisic (1997) hanno invece proposto una matrice dei pesi  $\mathbf{W}$  diagonale, i cui elementi sono costituiti dagli elementi diagonali della matrice di covarianza asintotica  $\hat{\Gamma}$ . La matrice dei pesi diagonale comporta una funzione di adattamento dei minimi quadrati ponderati diagonalmente. L'effetto positivo di questo approccio è che la matrice  $\mathbf{W}$  non deve essere invertita, operazione che può risultare problematica in modelli complessi, in campioni di piccole dimensioni o in presenza di variabili binarie particolarmente asimmetriche.

La metodologia proposta da Muthén, du Toit e Spisic (1997) ha molti meriti. Innanzitutto consente di analizzare un modello ad equazioni strutturali sufficientemente generale, anche se limitato a variabili risposta latenti distribuite normalmente. Questa metodologia è molto efficiente dal punto di vista computazionale, poiché riduce un problema che coinvolge integrali a più dimensioni ad un problema limitato ad una serie di integrali ad una o due dimensioni, aspetto non trascurabile quando le variabili latenti sono piuttosto numerose. Tale approccio è inoltre efficiente, in quanto produce stime molto prossime a quelle di massima verosimiglianza. Simulazioni Monte Carlo (Muthén e Kaplan, 1992) hanno dimostrato che tale metodo non si comporta in presenza di modelli piuttosto complessi se la numerosità campionaria è scarsa.

Muthén, du Toit e Spisic (1997) hanno infine realizzato uno studio Monte Carlo per confrontare la prestazione dello stimatore WLS robusto con quella dello stimatore GEE (Generalized Estimating Equation), proposto da Melton e Liang (1997) per la stima del modello ad equazioni strutturali di Muthén (1984) per variabili risposta binarie. Dallo studio emerge che lo stimatore WLS robusto presenta una variabilità inferiore a quella dello stimatore GEE ed è più rapido in termini computazionali. Il suo rendimento è invece inferiore in campioni di piccole dimensioni e con variabili risposta osservate asimmetriche.

Le ricerche empiriche sullo stimatore WLS robusto sono piuttosto limitate. I risultati dimostrano che tale stimatore è migliore rispetto allo stimatore WLS convenzionale quando il modello considerato è piuttosto grande (composto almeno da quindici variabili) o la numerosità campionaria è limitata (inferiore a 1000 unità osservate). Il suo rendimento è invece inferiore in campioni di piccole dimensioni e con variabili risposta osservate asimmetriche.



## 2.5 La standardizzazione dei parametri

Per facilitare l'interpretazione dei risultati, i parametri possono essere standardizzati. Il modello completo prevede due possibili strategie di standardizzazione. La prima riguarda solo i parametri del modello a variabili latenti e consente confronti limitati ai coefficienti che legano le variabili latenti. La seconda riguarda le variabili latenti e le variabili osservate e consente confronti tra tutti i coefficienti del modello. I parametri si possono standardizzare rispetto alla varianza delle variabili latenti  $\boldsymbol{\eta}$ , rispetto alla varianza delle variabili risposta osservate  $\mathbf{y}$  o rispetto alla varianza delle variabili esplicative  $\mathbf{x}$ .

La standardizzazione rispetto alle varianze unitarie delle variabili latenti  $\boldsymbol{\eta}$  si ottiene considerando la matrice diagonale  $\mathbf{D}_\eta$ , costituita degli scarti quadratici medi dei valori delle variabili latenti  $\boldsymbol{\eta}$ :

$$\mathbf{D}_\eta = \text{diag}\left[(\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Gamma} \boldsymbol{\Sigma}_{xx} \boldsymbol{\Gamma}' (\mathbf{I} - \mathbf{B})'^{-1} + (\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Psi} (\mathbf{I} - \mathbf{B})'^{-1}\right]^{1/2},$$

in modo che  $\boldsymbol{\eta}_i^* = \mathbf{D}_\eta^{-1} \boldsymbol{\eta}_i$  abbia varianza unitaria.

La standardizzazione rispetto alle varianze delle variabili di risposta latenti  $\mathbf{y}^*$  e quella rispetto alle variabili osservate  $\mathbf{x}$  si ottiene considerando rispettivamente le matrici diagonali  $\mathbf{D}_y$  e  $\mathbf{D}_x$ , la prima costituita dagli scarti quadratici medi delle  $\mathbf{y}^*$ , la seconda dagli scarti quadratici medi delle variabili  $\mathbf{x}$ . In particolare, la standardizzazione rispetto alle variabili risposta si ottiene nel modo seguente:

$$\mathbf{D}_y = \text{diag}\left[\boldsymbol{\Lambda}\left[(\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Gamma} \boldsymbol{\Sigma}_{xx} \boldsymbol{\Gamma}' (\mathbf{I} - \mathbf{B})'^{-1} \boldsymbol{\Lambda}' + (\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Psi} (\mathbf{I} - \mathbf{B})'^{-1} \boldsymbol{\Lambda}'^{-1} + \boldsymbol{\Theta}\right] \boldsymbol{\Lambda}\right]^{1/2},$$

mentre la standardizzazione rispetto alle variabili esplicative si ottiene considerando la matrice  $\mathbf{D}_x$ , i cui elementi sono la radice quadrata degli elementi diagonali di  $\boldsymbol{\Sigma}_{xx}$ . Infine, la standardizzazione si può fare anche rispetto allo scarto quadratico medio rispettivamente di  $\mathbf{y}^*$  ed  $\mathbf{x}$ . In questo caso, il coefficiente di determinazione  $R^2$  consente di misurare l'ammontare di variabilità di ciascuna variabile risposta spiegata

dalle espressioni standardizzate.

## **2.6 Adattamento del modello**

Una volta stimato il modello, è necessario valutare il suo adattamento ai dati. L'adattamento del modello ha due componenti primarie. La prima si riferisce alle misure di sintesi che valutano l'adattamento nel suo complesso mentre la seconda riguarda le misure che caratterizzano l'adattamento di singole componenti del modello.

Le misure di adattamento complessivo si distinguono tra statistiche test del modello e misure alternative di adattamento (o di adattamento approssimativo). Le statistiche test del modello misurano il grado di similarità tra la matrice di covarianza implicata dal modello e la matrice di covarianza campionaria. Tali statistiche si basano sulla distribuzione del chi-quadrato e restituiscono un livello di significatività associato al valore osservato della statistica utilizzata attraverso il quale è possibile respingere o non respingere l'ipotesi nulla sottoposta a verifica. Gli indici di adattamento alternativo sono misure continue di corrispondenza del modello ai dati e si possono ulteriormente distinguere<sup>59</sup> in indici di adattamento assoluto, indici di adattamento comparativo o incrementale, indici di parsimonia modificata e indici di adattamento predittivo. Trattandosi di misure continue di carattere descrittivo, un aspetto importante riguarda la possibilità di stabilire intervalli di valori che possono indicare un buon adattamento del modello ai dati. A questo proposito, Hu e Bentler (1999) hanno proposto una serie di valori soglia basandosi sui risultati ottenuti in alcuni studi di simulazione. In generale, tutte le misure di adattamento complessivo sono fondamentalmente caratterizzate da alcune limitazioni: i) indicano solo l'adattamento di un modello nel suo complesso, quindi è possibile che il modello abbia un pessimo adattamento in alcune sue parti nonostante l'adattamento complessivo sia buono; ii) riflettono solo un aspetto particolare dell'adattamento, perciò un valore favorevole di una misura non indica un adattamento accettabile; iii) esiste una relazione molto debole tra i valori delle misure di adattamento ed il grado o il tipo di mal specificazione di un modello (Millsap, 2007); iv) i valori delle misure di adattamento che suggeriscono un adattamento adeguato non sono in grado di stabilire il potere predittivo del modello, come nel caso di misure

---

<sup>59</sup> Queste categorie non sono mutuamente esclusive.

analoghe al R-quadrato; v) non sono in grado di indicare se i risultati sono significativi dal punto di vista teorico.

La considerazione di queste limitazioni porta a ritenere utile valutare anche l'adattamento di singole componenti del modello. Questo prevede la valutazione dell'esistenza di soluzioni improprie (varianze stimate negative e coefficienti di correlazione superiori ad uno in valore assoluto), la valutazione della plausibilità del segno e della grandezza dei coefficienti, delle varianze e delle covarianze, l'analisi del valore del coefficiente di determinazione riferito alle variabili endogene e la valutazione del senso alle stime dei parametri.

Quando le misure di adattamento del modello nel complesso o le misure di adattamento delle singole componenti sono giudicate non adeguate, allora è necessario specificare un nuovo modello.

#### 2.6.1 *La statistica test del modello*

La statistica test del modello consente di valutare la validità di un'ipotesi statistica. L'ipotesi nulla nell'analisi delle strutture di covarianza stabilisce l'uguaglianza tra la matrice di covarianza della popolazione e la matrice di covarianza implicata dal modello:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}) \text{ e } \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}).$$

In pratica, tale statistica consente di valutare se la matrice di covarianza implicata dal modello è simile alla matrice di covarianza campionaria e se le differenze osservate possono ritenersi dovute all'errore campionario. Tuttavia, poiché la matrice di covarianza ed i parametri della popolazione sono incogniti, si analizzano rispettivamente la matrice di covarianza campionaria  $\mathbf{S}$  e la matrice di covarianza implicata dal vettore dei parametri stimati,  $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ , ottenuta minimizzando una delle funzioni di adattamento  $F_{ML}$  o  $F_{GLS}$ .

Se l'ipotesi nulla è corretta, la funzione di adattamento valutata al suo valore minimo moltiplicata per la costante  $(N-1)$  converge asintoticamente ad una distribuzione chi-quadrato centrale con  $\frac{1}{2}(p+q)(p+q+3)-t$  gradi di libertà:

$$T = (N - 1)F(\mathbf{S}; \mathbf{\Sigma}(\hat{\boldsymbol{\theta}})) \sim \chi^2,$$

dove  $t$  è il numero di parametri incogniti. Quando il valore della statistica  $T = (N - 1)F$  non è significativo, l'unica cosa che si può affermare è che il modello è consistente con la matrice di covarianza ma non è possibile stabilire se il modello è corretto. Infatti, il modello potrebbe essere seriamente mal specificato ma avere un adattamento all'incirca equivalente a quello di uno dei tanti modelli equivalenti che implicano una matrice di covarianza identica o simile ai dati osservati (Hayduk, Cummings, Boadu, Pazderka-Robinson e Boulianne, 2007).

L'utilizzo della statistica  $T = (N - 1)F$  si basa sul rispetto delle seguenti assunzioni: i) variabili risposta osservate aventi una distribuzione normale multivariata; ii) l'analisi della matrice di covarianza; iii) un campione di numerosità sufficientemente grande; iv) un modello ipotizzato specificato in maniera corretta ed in grado di riflettere la reale struttura delle relazioni nella popolazione,  $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$ .

Di conseguenza, il valore osservato della statistica  $T = (N - 1)F$  può risultare influenzato da una serie di aspetti: i) il livello di non normalità multivariata delle variabili risposta osservate che può far aumentare o diminuire il suo valore con la conseguenza che l'adattamento del modello può risultare rispettivamente peggiore o migliore rispetto a quanto non sia in realtà (Hayduk, Cummings, Boadu, Pazderka-Robinson e Boulianne, 2007; Yuan, Bentler e Zhang, 2005); ii) la grandezza dei coefficienti di correlazione tra le variabili osservate che generalmente porta a valori maggiori della statistica chi-quadrato poiché favorisce il verificarsi di differenze maggiori tra le correlazioni (o covarianze) osservate e quelle previste dal modello; iii) la presenza di variabili con una proporzione elevata di variabilità specifica che provoca la perdita di potenza statistica; iv) la numerosità campionaria che tende a far aumentare il valore della statistica chi-quadrato (che può quindi diventare significativo) anche se le differenze tra le covarianze osservate e previste sono lievi.

### 2.6.2 Analisi dei residui

La matrice dei residui è forse la più semplice funzione della differenza tra  $\mathbf{S}$  e  $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$  che si può utilizzare per valutare l'adattamento complessivo di un modello. Gli elementi della matrice dei residui sono costituiti dalla differenza  $(s_{ij} - \hat{\sigma}_{ij})$ , dove  $s_{ij}$  rappresenta la covarianza campionaria tra le variabili  $i$  e  $j$  mentre  $\hat{\sigma}_{ij}$  il corrispondente valore implicato dal modello.

I residui sono influenzati da diversi fattori: i) dallo scarto tra  $\mathbf{\Sigma}$  e  $\mathbf{\Sigma}(\boldsymbol{\theta})$ ; ii) dalla scala delle variabili osservate; iii) dall'errore campionario. In particolare, poiché la grandezza dei residui è fortemente influenzata dall'unità di misura delle variabili, spesso si utilizzano le correlazioni residue:

$$r_{ij} - \hat{r}_{ij},$$

dove  $r_{ij}$  è il coefficiente di correlazione campionario tra le variabili  $i$  e  $j$  mentre  $\hat{r}_{ij}$  è il corrispondente valore implicato dal modello. La grandezza dei residui è influenzata dalla numerosità campionaria anche quando  $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$ . Per un dato modello  $(s_{ij} - \hat{\sigma}_{ij})$  tende a ridursi all'aumentare della numerosità campionaria.

Jöreskog e Sörbom (1986) hanno proposto un indice dei residui normalizzato che fornisce una correzione approssimativa rispetto alla numerosità campionaria:

$$NR = \frac{(s_{ij} - \hat{\sigma}_{ij})}{\left[ (\hat{\sigma}_{ii}\hat{\sigma}_{jj} + \hat{\sigma}_{ij}^2 / N) \right]^{1/2}},$$

dove la grandezza al denominatore rappresenta la radice quadrata della varianza asintotica stimata dei residui.

L'indice RMR (*Root Mean square Residual*) è stato proposto da Jöreskog e Sörbom (1981) come misura di adattamento complessivo basata sui residui:

$$RMR = \sqrt{\frac{\sum_{i=1}^p \sum_{j=1}^i (s_{ij} - \hat{\sigma}_{ij})^2}{p(p+1)/2}},$$

dove  $p$  è il numero di variabili. Valori prossimi a zero indicano un buon adattamento. Tuttavia, poiché gli elementi delle matrici  $\mathbf{S}$  e  $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$  dipendono dalla scala, anche i residui stimati dipendono dalla scala così e quindi l'indice RMR dipende dai valori di varianze e covarianze delle variabili osservate.

Per superare questo problema, è stato introdotto da Bentler (1995) l'indice SRMR (*Standardized Root Mean Square Residual*):

$$SRMR = \sqrt{\frac{\sum_{i=1}^p \sum_{j=1}^i \left( \frac{s_{ij} - \hat{\sigma}_{ij}}{\sqrt{s_{ii}} \sqrt{s_{jj}} - \sqrt{\hat{\sigma}_{ii}} \sqrt{\hat{\sigma}_{jj}}} \right)^2}{p(p+1)/2}},$$

in cui, rispetto all'indice RMR i residui sono divisi per le rispettive deviazioni standard. Anche in questo caso, valori dell'indice prossimi a zero indicano un buon adattamento: Tuttavia è difficile stabilire valori soglia perché l'indice è sensibile nel caso di modelli mal specificati e tende ad assumere valori minori quando la numerosità campionaria aumenta (Hu e Bentler, 1998). Nonostante ciò, si ritiene che l'adattamento sia buono quando l'indice assume valori inferiori a 0,05 mentre valori compresi tra 0,05 e 0,10 indicano un adattamento accettabile.

### 2.6.3 Misure di adattamento comparativo o incrementale

L'area di lavoro maggiormente attiva nell'ambito delle misure alternative è quella degli indici di adattamento comparativo o incrementale<sup>60</sup>. Tali indici misurano il miglioramento relativo dell'adattamento di un modello in confronto a quello di un modello di base, il cui adattamento è pessimo. Il modello di base è di solito quello di indipendenza che assume una matrice di covarianza nulla tra le variabili osservate,

---

<sup>60</sup> Per una discussione esaustiva sull'argomento è possibile consultare Hu e Bentler, 1995.

variabili osservate misurate senza errore (cioè con varianze degli errori fissate a zero) e coefficienti fattoriali pari a uno. Il modello di indipendenza prevede la stima solo di  $p$  parametri, pari al numero di varianze delle variabili osservate<sup>61</sup>. Un modello di base ancora più restrittivo è il modello nullo, in cui tutti i parametri sono fissati a zero (Jöreskog e Sörbom, 1993).

Un sottoinsieme di queste misure è stato progettato per tenere in considerazione il grado di mal specificazione del modello. In alcuni casi, questi indici sono arricchiti di una funzione di penalità per ciascun parametro stimato. In generale, tali indici sono normalizzati ed assumono valori compresi tra zero, adattamento pessimo, ed 1, adattamento perfetto.

L'indice di adattamento comparativo più rappresentativo è l'indice di adattamento normalizzato (*Normed Fit Index*) (Bentler & Bonnett, 1980):

$$NFI = \frac{\chi_b^2 - \chi_t^2}{\chi_b^2} = 1 - \frac{\chi_t^2}{\chi_b^2} = 1 - \frac{F_t}{F_b},$$

dove  $\chi_b^2$  è il chi-quadrato del modello di indipendenza completa<sup>62</sup> (baseline model),  $\chi_t^2$  è il chi-quadrato del modello considerato mentre  $F_t$  è il valore corrispondente al minimo della funzione di adattamento. Valori dell'indice prossimi a zero suggeriscono che il modello considerato non è molto migliore del modello di completa indipendenza mentre valori vicini ad uno<sup>63</sup> suggeriscono che il modello rappresenta un miglioramento significativo rispetto al modello di base. Di solito un valore pari a 0,95 è indicativo di un buon adattamento relativo rispetto al modello di base (Kaplan, 2000) anche se valori superiori a 0,90 sono tipicamente considerati indicativi di un adattamento accettabile (Marsh e Grayson, 1995; Schumacker e Lomaz, 1996). Uno svantaggio dell'indice NFI è che è influenzato dalla numerosità campionaria (Beraden, Sharma e Teel, 1982).

---

<sup>61</sup> Miles e Shelvin (2007) hanno notato che gli indici di adattamento comparativo basati sul modello di indipendenza effettivamente indicano il comportamento del modello considerato rispetto al “peggiore” modello che si possa considerare.

<sup>62</sup> Il modello di indipendenza completa è di solito associato ad un valore della statistica test del chi-quadrato elevato poiché l'ipotesi nulla testata da  $\chi_b^2$  stabilisce che non ci sia covarianza tra le variabili nella popolazione.

<sup>63</sup> Nonostante il limite teorico sia uno, l'indice NFI può non raggiungere questo limite superiore anche se il modello specificato è corretto, in particolare se la numerosità campionaria è piccola (Bentler, 1990).

Un indice analogo al precedente ma che introduce una penalità per la presenza di parametri aggiuntivi considerando i gradi di libertà del modello è l'indice di Tucker-Lewis (Tucker & Lewis, 1973), anche noto come indice di adattamento non normalizzato NNFI (*Non Normed Fit Index*):

$$TLI = NNFI = \frac{\chi_b^2/df_b - \chi_t^2/df_t}{\chi_b^2/df_b - 1},$$

dove  $df_b$  sono i gradi di libertà del modello di base (completa indipendenza) e  $df_t$  i gradi di libertà del modello considerato. Tale indice può produrre valori al di fuori dell'intervallo compreso tra 0 ed 1. Per un dato modello, un più basso rapporto tra chi-quadrato e gradi di libertà comporta un miglior adattamento del modello (purché tale rapporto non sia inferiore ad uno). Si noti, inoltre, che l'indice TLI dipende dall'ordine di grandezza media delle correlazioni osservate nei dati. Se queste non sono elevate, l'indice TLI non sarà molto elevato. L'indice TLI è uno dei meno influenzati dalla numerosità campionaria (Bentler, 1990; Bollen, 1990; Hu e Bentler, 1995, 1998).

Gli indici NFI e TLI assumono un'ipotesi nulla esatta e quindi una distribuzione centrale per la statistica test del chi-quadrato. Tuttavia, l'ipotesi nulla non è mai esattamente corretta e la distribuzione della statistica test può essere approssimata da una distribuzione chi-quadrato non centrale con parametro di non centralità  $\lambda$  (Kaplan, 2009). Una stima del parametro di non centralità può essere ottenuta come differenza tra la statistica test ed i suoi gradi di libertà. Perciò, per i modelli che non sono eccessivamente mal specificati, McDonald e Marsh (1990) hanno definito l'indice relativo di non centralità RNI (*Relative Noncentrality Index*):

$$RNI = \frac{(\chi_b^2 - df_b) - (\chi_t^2 - df_t)}{\chi_b^2 - df_b}.$$

Tale indice può assumere valori al di fuori dell'intervallo compreso tra 0 ed 1. Per rimediare a questo inconveniente, Bentler (1990) ha proposto una versione modificata dell'indice RNI, nota come indice CFI.



L'indice di adattamento comparativo CFI (*Comparative Fit Index*) consente di evitare la sottostima di adattamento osservata per l'indice NFI in campioni di piccole dimensioni. Nei modelli in cui  $\chi^2_t < df_t$ ,  $CFI = 1$ , altrimenti la formula è la seguente:

$$CFI = 1 - \frac{\chi^2_t - df_t}{\chi^2_b - df_b},$$

L'indice CFI varia tra zero ed uno con valori crescenti che indicano un adattamento migliore. Un valore unitario indica che  $\chi^2_t < df_t$  ma questo non significa perfetto adattamento del modello. Di solito, un valore di 0,97 è indicativo di un buon adattamento relativo rispetto al modello di indipendenza mentre valori maggiori di 0,95 sono considerati accettabili. Analogamente all'indice TLI-NNFI, l'indice CFI è uno degli indici meno influenzati dalla numerosità campionaria (Bentler, 1990; Bollen, 1990; Hu e Bentler, 1995, 1998).

L'indice GFI (*Goodness of Fit Index*) (Jöreskog e Sörbom, 1989; Tanaka e Huba, 1984) misura l'ammontare relativo della matrice di covarianza osservata  $S$  prevista dalla matrice di covarianza implicata dal modello,  $\Sigma(\hat{\theta})$ . L'indice GFI sembra ispirato al concetto di coefficiente di determinazione (Mulaik, James, Van Alstine, Bennett, Lind e Stillwell, 1989) ed è definito nel modo seguente:

$$GFI = 1 - \frac{\chi^2_t}{\chi^2_n} = 1 - \frac{F_t}{F_n},$$

dove  $\chi^2_n$  è il chi-quadrato del modello nullo (modello di base). L'indice GFI di solito assume valori compresi tra zero e uno con valori crescenti che indicano un adattamento migliore ma sono possibili anche valori negativi. Un valore di 0,95 è indicativo di un buon adattamento relativo rispetto al modello di base mentre valori maggiori di 0,90 sono ritenuti indicativi di un adattamento accettabile (Marsh e Grayson, 1995; Schumacker e Lomax, 1996).

Jöreskog e Sörbom (1989) hanno sviluppato l'indice AGFI (*Adjusted Goodness of Fit Index*) per aggiustare i gradi di libertà del modello rispetto al numero di variabili

osservate e quindi compensare (premiare) i modelli meno complessi con meno parametri. L'indice AGFI è dato da:

$$AGFI = 1 - \frac{df_b}{df_t} = (1 - GFI) = 1 - \frac{\chi_t^2/df_t}{\chi_b^2/df_b}.$$

L'utilizzo degli indici di adattamento comparativo non è privo di controversie. In particolare, Sobel e Borhnstedt (1985) sostengono che questi indici sono progettati per confrontare un modello ipotizzato contro una ipotesi piuttosto discutibile, rappresentata dal modello di base che specifica l'assenza di relazione tra le variabili osservate. Sobel e Borhnstedt affermano che tale ipotesi non è sostenibile e che tali indici dovrebbero utilizzare un modello di base più plausibile. Widaman e Thompson (2003) descrivono come specificare modelli di base più attendibili.

L'indice RMSEA (*Root Mean Square Error of Approximation*) è dato da:

$$RMSEA = \sqrt{\frac{\chi_t^2 - df_t}{df_t(N-1)}}.$$

Tale misura è scalata in maniera inversa rispetto agli altri indici e valori prossimi a zero indicano un buon adattamento. Quando  $\chi_t^2 \leq df_t$ , allora  $RMSEA = 0$  ma questo non implica necessariamente un adattamento perfetto perché questo non vuol dire  $\chi_t^2 = 0$ . Per valori  $\chi_t^2 > df_t$  il valore dell'indice RMSEA è progressivamente crescente. Considerato il denominatore, il valore dell'indice RMSEA diminuisce all'aumentare dei gradi di libertà e quindi della parsimonia del modello o all'aumentare della numerosità campionaria, a parità delle altre condizioni. Tuttavia, tale indice non favorisce necessariamente modelli con più gradi di libertà poiché l'effetto della correzione per la parsimonia diminuisce all'aumentare della numerosità campionaria (Mulaik, 2009).

#### 2.6.4 Criteri di selezione del modello

Un aspetto importante nella valutazione di un modello è la sua performance rispetto a quella di altri modelli concorrenti. In alcuni casi il confronto riguarda specificazioni

diverse di un modello stimato sullo stesso insieme di variabili osservate nel medesimo campione. Più spesso, i modelli concorrenti sono il risultato di modifiche non necessariamente gerarchiche di un modello iniziale. La scelta della misura da utilizzare per selezionare tra modelli concorrenti è diversa a seconda che i modelli siano gerarchici o non gerarchici. In generale, due modelli si dicono gerarchici se uno è un sottoinsieme dell'altro. In particolare, un modello (modello A) si dice compreso in un altro modello (modello B) avente più parametri e meno gradi di libertà, se il modello A può essere derivato dal modello B fissando<sup>64</sup> almeno un parametro incognito nel modello B o introducendo altre restrizioni. In pratica, due modelli sono compresi in una gerarchia quando i parametri incogniti del modello più restrittivo sono un sottoinsieme dei parametri incogniti del modello meno restrittivo.

Quando il confronto riguarda due modelli compresi in una gerarchia, si può utilizzare la statistica test differenza del chi-quadrato,  $\chi_D^2$ , per valutare la differenza di adattamento tra i due modelli concorrenti. Tale statistica si ottiene per differenza tra i valori chi-quadrato relativi ai due modelli gerarchici stimati sullo stesso insieme di dati (Steiger, Shapiro e Browne, 1985) ed ha per gradi di libertà la differenza tra i gradi di libertà dei due modelli,  $df_D$ :

$$\chi_D^2(df_D) = \chi_A^2(df_A) - \chi_B^2(df_B).$$

Se la differenza  $\chi_D^2$  è significativa, l'ipotesi nulla di uguale adattamento tra i due modelli viene respinta ed il modello B viene mantenuto. La statistica  $\chi_D^2$  ha essenzialmente gli stessi limiti della statistica  $\chi^2$  applicata ad un modello singolo.

Quando il confronto riguarda due modelli non compresi in una gerarchia, la statistica test differenza del chi-quadrato  $\chi_D^2$  può essere valutata ma il suo valore non può essere interpretato come una statistica test. In questo caso, per selezionare tra modelli concorrenti si possono utilizzare altri indici, come il *Akaike Information Criteria* (AIC) e il *Bayesian Information Criterion* (BIC).

L'indice AIC (*Akaike Information Criteria*) è stato sviluppato da Akaike (1973, 1987) ed è basato sulla teoria dell'informazione combina stima e selezione del modello

---

<sup>64</sup> Di solito, il valore del parametro viene fissato a zero e corrisponde ad eliminare il parametro.

in un unico ambito concettuale (Anderson, Burnham e Thompson, 2000). Tale indice è il più noto nell'ambito della stima ML e si può anche considerare un indice di parsimonia modificato perché tende a favorire modelli più semplici. In letteratura sono presenti diverse versioni (Hayduk, 1996), due di queste sono riportate di seguito. La prima formula è data da:

$$AIC_1 = \chi^2_t + 2q ,$$

dove  $q$  è il numero di parametri incogniti del modello. Da questa si evince che il valore della statistica chi-quadrato viene aumentata di un fattore pari a due volte il numero di parametri da stimare. La seconda formula è data da:

$$AIC_2 = \chi^2_t - 2df_t ,$$

dalla quale si ricava che il valore della statistica chi-quadrato viene diminuito di un fattore pari a due volte il numero di gradi di libertà. Nonostante le due formule siano diverse, il punto essenziale è che la variazione relativa dell'indice AIC è la stessa per entrambe le versioni della formula in quanto è funzione della complessità del modello. In particolare, si può notare che la correzione relativa per la parsimonia del modello diviene sempre più piccola all'aumentare delle numerosità campionaria (Mulaik, 2009). Nella scelta tra modelli concorrenti, il modello che presenta un indice AIC più basso è considerato migliore. In pratica, questo rappresenta il modello che presenta l'adattamento relativamente migliore e un numero di parametri inferiore rispetto ai modelli concorrenti. Tuttavia, questo indice non avendo una statistica campionaria associata, non è in grado di stabilire se il modello si adatta bene ai dati.

L'indice BIC (*Bayesian Information Criterion*) è stato proposto da Schwartz (1978) e Raftery (1986, 1993) e si basa sulla teoria bayesiana di selezione di un modello:

$$BIC = -2 \ln L + q_0 \ln(n) ,$$

dove  $q_0$  è il numero di parametri secondo l'ipotesi nulla e  $n$  la numerosità campionaria. Tale indice incrementa la penalità per il numero di parametri stimati in funzione del logaritmo della numerosità campionaria<sup>65</sup>. In generale, L'indice BIC tende a penalizzare modello con troppi parametri in maniera più accentuata di quanto avvenga con l'indice AIC.

### *2.6.5 Effetti con variabili osservate non distribuite normalmente o aventi modalità categoriche*

Le conseguenze sulle misure di adattamento complessivo (statistiche test del modello e misure alternative di adattamento) sono diverse a seconda che le variabili osservate abbiano una distribuzione non normale o modalità categoriche.

Quando le variabili osservate hanno distribuzioni che si differenziano da quella normale, la statistica  $\chi^2$  risulta sovrastimata anche in presenza di scostamenti moderati dalla forma normale con valori che tendono ad aumentare al crescere del grado di non normalità (Chou, Bentler e Satorra, 1991; Hu, Bentler e Kano, 1992; Curran, West e Finch, 1996; Yu e Muthén, 2002). L'effetto maggiore si verifica nel caso di distribuzioni leptocurtiche (Browne, 1984; Chou, Bentler e Satorra, 1991). Ciò comporta errori di I tipo più frequenti<sup>66</sup>. Gli indici di adattamento espressi in funzione del  $\chi^2$  (come ad esempio gli indici di adattamento comparativo) hanno un comportamento analogo, poiché sono influenzati dagli stessi fattori (Hu e Bentler, 1998). Di conseguenza, in presenza di distribuzioni non normali, gli indici TLI, CFI ed RMSEA calcolati nel caso di stimatori ML tendono a respingere modelli specificati in maniera corretta (Hu e Bentler, 1999; Yu e Muthén, 2002).

Quando le variabili presentano modalità categoriche, gli indici di adattamento mostrano un buon rendimento se la forma della distribuzione è approssimativamente normale e le variabili hanno almeno cinque modalità (Babakus, Ferguson e Jöreskog, 1987; Hutchinson e Holmos, 1998). L'indice  $\chi^2$  è robusto anche con variabili categoriche che presentano solo quattro modalità ma il suo valore tende a crescere quando il numero di modalità è inferiore (Green, Akey, Fleming, Hershberger e

---

<sup>65</sup> Considerando una funzione logaritmica, la penalità tende ad aumentare con tassi di crescita decrescenti all'aumentare della numerosità campionaria.

<sup>66</sup> Questo comporta un più alto tasso di rifiuto di modelli specificati in maniera corretta.

Marquis, 1997). Invece, gli indici GFI, AGFI e RMR sono leggermente sottostimati se la dimensione campionaria è piccola ed il numero di modalità è pari a cinque. In definitiva, analizzare i dati come se fossero continui comporta un livello di distorsione non significativo degli indici di adattamento in presenza di variabili aventi una distribuzione approssimativamente normale con almeno cinque modalità ordinate (Muthén e Kaplan, 1985; Bollen, 1989; Dolan, 1994). Quando la distribuzione non è normale, i valori dell'indice  $\chi^2$  basato sulla funzione ML (Green, Akey, Fleming, Hershberger e Marquis, 1997; West, Finch e Curran, 1995) e dell'indice RMR tendono a crescere ed i valori degli indici NNFI, GFI e CFI risultano sottostimati (Babakus, Ferguson e Jöreskog, 1987; Hutchinson e Holmos, 1998). Questo comporta che un modello specificato correttamente non si adatta bene ai dati e può essere respinto.

#### 2.6.6 Misure di adattamento con variabili non distribuite normalmente

Nei casi in cui l'ipotesi di normalità non è rispettata, un'espressione più generale per il test chi-quadrato di adattamento del modello è stata fornita da Browne (1984). Si consideri la seguente forma quadratica basata sui residui (Browne, 1984; Satorra e Bentler, 1990, 1994; Satorra, 1990, 1992):

$$nT = (s - \hat{\sigma})' \mathbf{A} (s - \hat{\sigma}),$$

dove  $\hat{\sigma} = (\hat{\boldsymbol{\theta}})$  mentre  $\mathbf{A}$  è uno stimatore consistente di:

$$\mathbf{W} - \mathbf{W}\mathbf{A}(\mathbf{A}'\mathbf{W}\mathbf{A})^{-1}\mathbf{A}'\mathbf{W} = \mathbf{A} \perp (\mathbf{A} \perp' \mathbf{W}^{-1} \mathbf{A} \perp)^{-1} \mathbf{A} \perp',$$

dove  $\mathbf{A} \perp$  è un complemento ortogonale di  $\mathbf{A}$  (Satorra, 1992). La robustezza in assenza di normalità si ottiene considerando  $\mathbf{W} = \hat{\boldsymbol{\Gamma}}^{-1}$ . Satorra (1989) ha dimostrato che questo risultato si mantiene anche quando il minimo della funzione di adattamento è ottenuto facendo riferimento alla teoria normale.

Un approccio alternativo per valutare l'adattamento del modello quando la forma della distribuzione delle variabili osservate è mal specificata è stato proposto da Satorra e Bentler (1988, 1994) nell'ambito dell'analisi delle strutture di covarianza e

prevede di modificare l'indice  $\chi^2$  con un fattore di scala che tiene conto del grado di non normalità nei dati. Tale indice di adattamento robusto è stato definito come *mean-adjusted chi-square* ed assume la forma seguente:

$$T_1 = \frac{d}{\text{tr}[(\mathbf{U}\mathbf{\Gamma})]} T,$$

dove  $T = (N-1)F$ ,  $\mathbf{U} = (\mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{\Delta}(\mathbf{\Delta}'\mathbf{W}^{-1}\mathbf{\Delta})^{-1}\mathbf{\Delta}'\mathbf{W}^{-1})$  e  $d$  sono i gradi di libertà del modello.

Una misura che tiene conto anche della varianza è invece il *mean and variance adjusted chi-square* (Satterthwaite, 1941; Satorra e Bentler, 1988, 1994):

$$T_2 = \frac{d}{\text{tr}(\mathbf{U}\mathbf{\Gamma})^2} T,$$

dove  $d$  è calcolato come il valore intero più vicino a  $d^* = (\text{tr}(\mathbf{U}\mathbf{\Gamma}))^2 / \text{tr}((\mathbf{U}\mathbf{\Gamma})^2)$ . Da notare, anche in questo caso, che sia  $T_1$  che  $T_2$  non richiedono l'inversione della matrice  $\mathbf{\Gamma}$  ma solo di  $\mathbf{W}$ .

Muthén (1993) ha realizzato uno studio di simulazione Monte Carlo in cui mostra che la prestazione dell'indice di adattamento robusto  $T_1$  è migliore del tradizionale  $T$ .

L'utilizzo dello stimatore WLS robusto si può accompagnare all'indice WRMR (*Weighted Root Mean square Residual*), definito nel modo seguente:

$$WRMR = \sqrt{\frac{\sum_r^e \frac{(s_r - \hat{\sigma}_r)^2}{v_r}}{e}},$$

dove  $s_r$  è un elemento del vettore delle statistiche campionarie,  $\hat{\sigma}_r$  è il corrispondente valore stimato,  $v_r$  è una stima della varianza asintotica di  $s_r$  ed  $e$  è il numero di statistiche campionarie. In generale, l'indice WRMR è adatto nelle situazioni in cui le

variabili risposta osservate presentano varianze piuttosto diverse e quando le variabili risposta osservate sono misurate su scale diverse<sup>67</sup> poiché incorpora le varianze asintotiche (Muthén e Muthén, 2001). L'indice WRMR è adatto anche in presenza di variabili risposta continue non distribuite normalmente o di variabili risposta categoriche. Valori dell'indice inferiori a 0,90 rappresentano un buon adattamento del modello mentre valori più piccoli indicano un adattamento migliore (Yu e Muthén, 2002).

---

<sup>67</sup> Un esempio di questo tipo si ha nel caso di modelli che presentano una struttura parametrica per la media e per le soglie.



## 2.7 L'analisi multilivello

L'analisi multilivello è una metodologia statistica che consente di analizzare in un unico modello statistico relazioni tra variabili riferite ad unità statistiche di livello<sup>68</sup> diverso. I modelli statistici multilivello consentono di spiegare strutture di variabilità complesse ed analizzano la variabilità associata a ciascun livello. Tali modelli sono noti nei diversi ambiti di ricerca anche come modelli gerarchici (Lindley e Smith, 1972; Raudenbush e Bryk, 1986), modelli ad effetti misti, modelli ad effetti casuali (de Leeuw e Kreft, 1986; Longford, 1993) o modelli a componenti di varianza (Longford, 1987; Searle, Casella e McCulloch, 1992).

I dati per l'analisi multilivello hanno di solito una struttura gerarchica<sup>69</sup> che si realizza quando le unità di primo livello (unità elementari o individuali) sono comprese in unità di livello superiore ma possono riferirsi anche ad altre tipologie di strutture (Gelman e Hill, 2007), come accade per i dati derivanti da una classificazione incrociata<sup>70</sup> (Snijders e Bosker, 1999; Raudenbush e Bryk, 2002; Goldstein, 2003) o per strutture ad appartenenza multipla<sup>71</sup> (Leyland e Goldstein, 2001; Fielding e Goldstein, 2006). La variabile risposta è misurata al livello inferiore della struttura multilivello mentre le variabili esplicative sono potenzialmente misurate a tutti i livelli presenti.

I dati aventi una struttura gerarchica si possono ottenere anche in seguito all'osservazione di risposte ripetute<sup>72</sup>, di risposte multivariate<sup>73</sup>, a studi di meta analisi<sup>74</sup>, o ad una procedura di campionamento a grappoli<sup>75</sup> o a stadi<sup>76</sup>. La conseguenza di questa

---

<sup>68</sup> In generale, il termine livello si riferisce alla posizione di una unità statistica all'interno della struttura di un modello.

<sup>69</sup> Nel caso di dati aventi una struttura gerarchica, il termine livello si usa per indicare la posizione di una unità statistica all'interno di una gerarchia di unità in cui le unità statistiche al livello più basso della gerarchia (livello 1) sono comprese in unità di livello superiore (livello 2) e così via fino al livello più elevato considerato (livello k).

<sup>70</sup> In presenza di dati derivanti da una classificazione incrociata, le unità sono classificate in maniera congiunta secondo due o più fattori e possono appartenere potenzialmente ad una sola delle possibili combinazioni di modalità dei diversi fattori. In questo caso le combinazioni di modalità rappresentano le unità di livello 2 all'interno delle quali si trovano le unità di livello 1.

<sup>71</sup> In presenza di strutture ad appartenenza multipla, le unità sono comprese in più di una unità di livello superiore. Anche in questo caso le combinazioni di modalità dei diversi fattori rappresentano le unità di livello 2 all'interno delle quali si trovano le unità di livello 1.

<sup>72</sup> In presenza di risposte ripetute nel tempo, le unità statistiche sono osservate in più occasioni. In questo caso le occasioni rappresentano le unità di livello 1 e sono comprese nelle unità di livello 2 che sono rappresentate dalle unità osservate.

<sup>73</sup> In presenza di risposte multivariate, si osservano diverse risposte per ciascuna unità statistica. In questo caso le risposte rappresentano le unità di livello 1 e sono comprese nelle unità di livello 2 che rappresentano le unità osservate.

<sup>74</sup> In presenza di studi meta analisi, le unità sono comprese in studi diversi.

<sup>75</sup> Il campionamento a grappoli prevede la selezione casuale di un certo numero di gruppi e,

struttura nei dati è che le unità statistiche appartenenti allo stesso gruppo per ciascun livello della gerarchia, se condividono fattori non osservati, possono avere valori della variabile risposta correlati, anche dopo il condizionamento alle variabili esplicative. In questo caso le osservazioni non sono indipendenti ed identicamente distribuite ed i metodi statistici tradizionali non sono adeguati. Infatti, quando l'assunzione di indipendenza non è rispettata (nell'analisi multilivello questa è una situazione frequente) le stime degli errori standard sono troppo piccole ed i parametri possono risultare significativi in maniera spuria. L'obiettivo dell'analisi multilivello è quello di determinare gli effetti delle variabili esplicative osservate ai diversi livelli sulla variabile risposta e di stabilire se le variabili esplicative a livello superiore agiscono da moderatori per le relazioni a livello inferiore<sup>77</sup>.

### *2.7.1 Aggregazione e disaggregazione delle variabili*

Nell'analisi multilivello le variabili si possono definire per ciascun livello della struttura e possono essere di tipo diverso<sup>78</sup>. Alcune di queste sono misurate al loro livello "naturale" e sono le variabili direttamente riferite alle unità che descrivono<sup>79</sup>, altre si possono ricavare spostando variabili da un livello ad un altro della struttura per aggregazione o disaggregazione di variabili osservate a livelli diversi<sup>80</sup>. L'aggregazione consente di assegnare alle unità di livello superiore della struttura uno stesso valore, ricavato da una sintesi di variabili appartenenti a livello inferiore. La disaggregazione consente di assegnare alle unità di livello inferiore della struttura i valori assunti da una variabile osservata a livello superiore. Le variabili che risultano dalla disaggregazione si

---

successivamente, l'inclusione nel campione di tutte le unità appartenenti a ciascun gruppo.

<sup>76</sup> Il campionamento a stadi prevede la selezione causale di un certo numero di gruppi, e successivamente, l'estrazione di un certo numero di unità dai gruppi selezionati. Nel campionamento a stadi, la procedura può essere naturalmente reiterata per un numero arbitrario di stadi.

<sup>77</sup> Quando le variabili osservate a livello superiore esercitano un effetto moderatore sulle relazioni a livello inferiore, si manifesta una interazione statistica tra variabili esplicative osservate a livelli diversi.

<sup>78</sup> Per una descrizione dettagliata della relazione tra i diversi tipi di variabili, definite a livelli differenti, si può fare riferimento a Lazarsfeld e Menzel (1961).

<sup>79</sup> Tali variabili sono state distinte tra globali e relazionali. Le variabili globali si riferiscono soltanto alle unità del livello per il quale sono definite, senza alcun riferimento alle altre unità o livelli. Le variabili relazionali, pur appartenendo al livello per il quale sono definite, descrivono la relazione di una unità con le altre unità allo stesso livello della struttura.

<sup>80</sup> Tali variabili sono state distinte tra analitiche e strutturali e sono riferite alle unità di livello inferiore della struttura. Le variabili analitiche sono costruite come sintesi di variabili a livello inferiore. Le variabili strutturali si riferiscono invece alla distribuzione delle variabili relazionali a livello inferiore.

definiscono anche variabili di contesto poiché caratterizzano le unità di livello superiore, cioè l'”ambiente” in cui si trovano le unità di livello inferiore<sup>81</sup>.

Inizialmente, l'analisi dei dati aventi una struttura multilivello è stata fatta utilizzando i metodi statistici tradizionali dopo aver ricondotto le unità ad un unico livello, aggregando a livello di gruppo le variabili individuali o disaggregando a livello individuale le variabili riferite a livello di gruppo. Tuttavia, analizzare variabili osservate a livelli diversi come se fossero appartenenti ad uno stesso livello, ignorando la struttura multilivello, può portare a due ordini di problemi.

Il primo problema è di tipo statistico e deriva dalle procedure di aggregazione o disaggregazione. Quando i dati sono aggregati si perde gran parte dell'informazione o di variabilità contenuta nelle unità di livello inferiore poiché i valori associati a ciascuna unità sono raggruppati in un numero minore di valori, corrispondenti alle unità di livello superiore. Il risultato è che l'analisi statistica perde in potenza e la relazione che si vuole studiare può apparire più forte di quanto non sia in realtà. Invece, quando i dati sono disaggregati, i valori riferiti alle unità di livello superiore sono esplosi in un numero maggiore di valori, corrispondenti alle unità osservate a livello inferiore che vengono così ad assumere lo stesso valore delle unità di livello superiore in cui sono comprese. In questo caso l'assunzione di indipendenza degli errori è rispettata solo per le unità di livello superiore ed i metodi statistici tradizionali, basati sull'ipotesi di indipendenza tra le unità a livello elementare, non sono appropriati. Inoltre, per le variabili risultanti dalla procedura di disaggregazione, la numerosità campionaria corretta è il numero di unità di livello superiore. In questo caso, utilizzare una numerosità campionaria pari al numero di unità di livello elementare, porta a rifiutare l'ipotesi nulla molto più spesso di quanto stabilito dal livello di significatività<sup>82</sup> poiché gli errori standard risultano sottostimati (Hox, 2002).

Oltre a questi problemi, si deve aggiungere che l'aggregazione non consente di studiare le relazioni tra livelli, elemento essenziale dell'analisi multilivello mentre la disaggregazione non consente di fare inferenza sulle unità di livello superiore<sup>83</sup>.

---

<sup>81</sup> Per le variabili di contesto ottenute come medie disaggregate di gruppo, Cronbach (1976) ha suggerito di esprimere i punteggi (valori) a livello individuale come scarti dalla rispettiva media di gruppo, procedura nota come centratura rispetto alla media di gruppo. In questo caso i punteggi individuali devono essere interpretati in termini relativi rispetto alla media di gruppo.

<sup>82</sup> Questo vuol dire che si possono manifestare falsi risultati significativi.

<sup>83</sup> Questo significa che non si possono trattare i gruppi osservati come un campione casuale estratto da

Il secondo ordine di problemi che si può avere ignorando la struttura multilivello dei dati è di tipo concettuale. Se non si presta attenzione all'interpretazione dei risultati, si può commettere l'errore di analizzare dati riferiti ad unità appartenenti ad un livello e formulare conclusioni riferite ad unità di un altro livello. In questo caso, l'errore più comune è la fallacia ecologica o "effetto di Robinson" (Robinson, 1950; Kreft, de Leeuw, 1987) e si può verificare quando si interpretano a livello individuale i risultati di analisi di dati raccolti a livello aggregato. Il problema opposto è noto invece come "fallacia atomistica" e si può verificare quando i risultati sono interpretati a livello aggregato ma i dati sono raccolti a livello individuale (Alker, 1969). Un errore leggermente differente è il paradosso di Simpson (Lindley e Novick, 1981) e si può verificare quando si ignora l'eterogeneità in una popolazione e si raggiungono conclusioni completamente errate perché i dati sono analizzati come se fossero provenienti da una singola popolazione omogenea.

Tali modelli assumono una struttura dei dati gerarchica, con una singola variabile risposta misurata al livello inferiore e variabili esplicative potenzialmente misurate su tutti i livelli esistenti nella gerarchia.

### *2.7.2 Prospettive di analisi*

L'analisi dei dati aventi una struttura multilivello è un argomento piuttosto complesso che attinge da diverse aree di ricerca metodologica ed è stato affrontato anche nell'ambito dello studio dei metodi statistici per l'analisi dei dati raccolti con disegni di campionamento complessi<sup>84</sup> (Asparouhov e Muthén, 2006; Rabe-Hesketh e Skrondal, 2008).

Muthén e Satorra (1995) hanno discusso due possibili approcci per l'analisi dei modelli ad equazioni strutturali in presenza di una struttura dati ottenuta in base ad una procedura di campionamento complessa, distinguendo tra analisi aggregata ed analisi disaggregata. L'analisi aggregata definisce i parametri di un modello ad equazioni strutturali convenzionale mentre gli errori standard ed i test di adattamento sono modificati per tenere conto della struttura dei dati. Questo approccio è consigliabile

---

una popolazione di gruppi.

<sup>84</sup> Tali disegni comprendono il campionamento stratificato, a stadi, a grappoli e possono essere caratterizzati anche dalla presenza di probabilità di selezione ineguale delle unità, stratificazione a posteriori e mancate risposte. Per una trattazione esaustiva dell'argomento si può consultare Skinner, Holt e Smith (1989) e Longford (1995).

quando si è interessati ad analizzare le relazioni tra variabili osservate a livello individuale e la struttura multilivello dei dati costituisce solo un fattore di disturbo. L'analisi disaggregata definisce invece un modello più complesso che prevede un insieme di nuovi parametri che riflettono la struttura complessa dei dati. Questo approccio assume un obiettivo più ambizioso poiché cerca di spiegare anche l'eterogeneità osservata nella popolazione ed è preferibile quando la struttura multilivello dei dati costituisce l'interesse primario dell'analisi.

Le conseguenze sugli errori standard delle stime quando si assume un campionamento casuale semplice in presenza di dati ottenuti in seguito ad una procedura di campionamento a grappoli o a stadi sono abitualmente studiate in termini di effetti del disegno (Kish, 1965; 1987). L'effetto del disegno è una quantità che indica di quanto si devono modificare gli errori standard in conseguenza alla strategia di campionamento adottata. In un disegno a due livelli, l'effetto del disegno si ottiene nel modo seguente (Cochran, 1977)<sup>85</sup>:

$$deff = \frac{V_c}{V_{srs}} = 1 + (n_{clus} - 1)\rho,$$

dove  $V_c$  è la varianza dello stimatore in condizioni di campionamento a stadi,  $V_{srs}$  è la varianza dello stimatore in caso di campionamento casuale semplice,  $n_c$  è la numerosità costante dei gruppi e  $\rho$  il coefficiente di correlazione intraclasse<sup>86</sup>. Il coefficiente di correlazione intraclasse misura la proporzione di variabilità della variabile osservata spiegata dalla variabile di raggruppamento (unità di secondo livello) e si calcola nel modo seguente:

$$\rho = \frac{MS_c - MS_w}{MS_c + (df_c)MS_w},$$

dove  $MS_c$  e  $MS_w$  rappresentano rispettivamente la somma dei quadrati tra i gruppi ed

---

<sup>85</sup> La formula riportata di seguito si applica al campionamento a due stadi. Altri schemi di campionamento, come ad esempio quello stratificato, richiedono formule diverse (Kish, 1965, 1987).

<sup>86</sup> Il simbolo  $\rho$  è stato introdotto da Kish (1965) che lo definì *roh*, "rate of homogeneity".

entro i gruppi mentre  $df_c$  sono i gradi di libertà tra i gruppi. Nell'ipotesi più realistica di gruppi non bilanciati, la numerosità campionaria effettiva si ottiene nel modo seguente:

$$n_c = \frac{N^2 - \sum_{g=1}^G n_g^2}{N(G-1)},$$

dove  $N$  è la numerosità campionaria totale e  $G$  il numero di gruppi.

L'effetto del disegno evidenzia che quando si assume in modo improprio un campionamento casuale semplice, il livello di sottostima degli errori standard è dovuto all'effetto combinato della dimensione dei gruppi,  $n_c$ , e del coefficiente di correlazione intraclasse,  $\rho$ . Quando  $\rho = 0$  l'effetto del disegno assume valore unitario e le unità sono indipendenti anche per ciascun livello di raggruppamento. Maggiore è il coefficiente di correlazione intraclasse, maggiore è lo scostamento dall'assunzione di indipendenza e quindi la distorsione che si ottiene considerando le procedure inferenziali basate sull'ipotesi di indipendenza. In generale, gli errori standard ottenuti in caso di campionamento casuale semplice risultano sottostimati non appena il coefficiente di correlazione intraclasse assume valori positivi.

## 2.8 Modelli ad equazioni strutturali multilivello

I modelli ad equazioni strutturali considerati finora assumono osservazioni indipendenti ed identicamente distribuite. Tale assunzione non è rispettata quando le unità statistiche elementari sono comprese in una struttura gerarchica o, più in generale, quando sono comprese in una struttura multilivello. In questo caso, utilizzare i modelli ad equazioni strutturali convenzionali assumendo un campionamento casuale semplice delle unità osservate produce stime dei coefficienti distorte (Muthén e Satorra, 1989).

L'approccio tradizionale per estendere i modelli ad equazioni strutturali al caso multilivello quando le variabili risposta osservate sono continue o categoriche modellate da variabili continue latenti sottostanti è quello di formulare modelli separati per le matrici di covarianza entro i gruppi (*within-groups*) e tra i gruppi (*between-groups*). Tale approccio è riconducibile a Goldstein e McDonald (1988) e McDonald e Goldstein

(1989)<sup>87</sup> ma è stato successivamente affiancato da alcune formulazioni non completamente equivalenti ma piuttosto simili tra loro (Muthén, 1989 e 1994; Muthén e Satorra, 1989 e 1995; Lee, 1990; Raudenbush, 1995; Lee e Poon, 1998; Lee e Shi, 2001; Lee e Song, 2004; du Toit e du Toit, 2004; Liang e Bentler, 2004). Quasi tutti questi modelli prevedono stimatori di massima verosimiglianza<sup>88</sup> con variabili osservate continue<sup>89</sup>. Asparouhov e Muthén (2007) hanno proposto un modello ad equazioni strutturali a due livelli con variabili risposta osservate binarie, categoriche ordinali, censurate, continue modellate da variabili continue latenti sottostanti. Tale approccio, prevede la formulazione di modelli separati per le matrici di covarianza entro i gruppi e tra i gruppi ma si differenzia dai precedenti perché si basa sul metodo di stima ad informazione limitata e sulla funzione di adattamento dei minimi quadrati ponderati (Muthén, 1984; Muthén, du Toit, e Spisic, 1997). Il vantaggio dello stimatore dei minimi quadrati ponderati è che si può utilizzare per stimare modelli con un numero qualsiasi di effetti casuali senza che questo comporti un allungamento dei tempi di calcolo o comprometta la precisione delle stime.

Rabe-Hesketh, Skrondal e Pickles (2004) hanno proposto un approccio per l'analisi dei modelli ad equazioni strutturali generalizzati multilivello che non prevede la stima di matrici di covarianza separate tra i gruppi ed entro i gruppi. Tale approccio rientra nell'ambito dei modelli lineari generalizzati misti a variabili latenti (GLLAMM)<sup>90</sup>, consente di specificare due o più livelli di analisi ed è basato su uno stimatore di massima verosimiglianza<sup>91</sup>. Il modello è composto da un modello di risposta (o modello di misura) e da un modello a variabili latenti. Il modello di misura è un modello lineare generalizzato misto e permette di modellare un'ampia gamma di

---

<sup>87</sup> Uno dei primi tentativi di sviluppo dell'analisi a due livelli con variabili latenti è invece dovuto a Schmidt (1969).

<sup>88</sup> Muthén (1989) ha presentato uno stimatore (MUML) che restituisce stime di massima verosimiglianza con dati bilanciati e una soluzione approssimata con dati non bilanciati. Lee (1990) ha proposto invece uno stimatore basato sulla funzione di adattamento del metodo dei minimi quadrati generalizzati.

<sup>89</sup> I modelli di Lee e Shi (2001) e Lee e Song (2004) considerano invece variabili risposta categoriche modellate da variabili continue latenti sottostanti.

<sup>90</sup> L'acronimo GLLAMM significa *Generalized Linear Latent And Mixed Models*. Lo sviluppo dell'ambiente GLLAMM è avvenuto in parallelo a quello del programma statistico "gllamm" che lavora in Stata, ed è disponibile per il download al seguente indirizzo internet: [www.gllamm.org](http://www.gllamm.org). Tale programma può stimare tutti i modelli previsti dall'ambiente GLLAMM ad eccezione, per il momento, dei modelli che prevedono variabili latenti discrete e continue. Inoltre, per le variabili latenti continue l'unica distribuzione disponibile è la distribuzione normale multivariata.

<sup>91</sup> In presenza di variabili risposta non continue, non esiste una soluzione in forma chiusa per la funzione di verosimiglianza marginale. In questo caso Rabe-Hesketh, Skrondal e Pickles (2004) suggeriscono di utilizzare l'algoritmo Newton-Raphson in cui le variabili latenti sono integrate con la quadratura adattiva.

variabili risposta.

In questo paragrafo è descritta la specificazione e la stima dell'analisi disaggregata dei modelli ad equazioni strutturali. Dopo aver esposto le caratteristiche dell'analisi disaggregata ed il modello di misura a due livelli (paragrafi 2.8.1 e 2.8.2), sono descritti in maniera dettagliata l'approccio pseudobilanciato di Muthén (*pseudobalanced Muthén's approach*) (paragrafo 2.8.3) ed il metodo basato sui minimi quadrati ponderati ad informazione limitata (Asparouhov e Muthén, 2007) (paragrafo 2.8.6) mentre i principali contributi basati sugli stimatori di massima verosimiglianza sono brevemente ricordati (paragrafi 2.8.4).

### 2.8.1 L'analisi disaggregata a due livelli

Il punto di partenza è una struttura dati a due livelli in cui sono osservate  $N$  unità statistiche suddivise in  $J$  gruppi<sup>92</sup> per la quale si assume variabilità nelle risposte a livello individuale e a livello di gruppo. In questo caso, il vettore di variabili osservate riferite all'unità  $i$  appartenente al gruppo  $j$ , indicato con  $\mathbf{y}_{ij}$ , si può scomporre in una componente che varia tra i gruppi,  $\mathbf{y}_B = \bar{\mathbf{y}}_j$ , e una componente che varia entro i gruppi,  $\mathbf{y}_W = \mathbf{y}_{ij} - \bar{\mathbf{y}}_j$  (Cronbach e Webb, 1979). In pratica, per ciascuna unità si sostituisce il vettore di variabili osservate,  $\mathbf{y}_T = \mathbf{y}_{ij}$ , con le sue componenti, la componente di gruppo,  $\mathbf{y}_B$ , e la componente individuale,  $\mathbf{y}_W$ . Queste due componenti sono ortogonali e additive (Searle, Casella e McCulloch, 1992):

$$\mathbf{y}_T = \mathbf{y}_B + \mathbf{y}_W.$$

La scomposizione del vettore delle variabili osservate può essere utilizzata per calcolare una matrice di covarianza tra i gruppi, indicata con  $\Sigma_B$ , e una matrice di covarianza entro i gruppi, indicata con  $\Sigma_W$ . Tali matrici rappresentano, rispettivamente, la matrice di covarianza riferita agli scarti tra le medie di gruppo e la media generale e la matrice di covarianza riferita agli scarti delle unità dalle medie di ciascun gruppo e sono a loro volta ortogonali e additive:

---

<sup>92</sup> Si assume, inoltre, che il campione osservato di gruppi sia stato selezionato in maniera casuale da una popolazione di gruppi e che le unità siano state selezionate casualmente all'interno di ciascun gruppo.



$$\Sigma_T = \Sigma_B + \Sigma_W .$$

In maniera analoga è possibile scomporre anche le osservazioni campionarie, rispetto alle quali anche le matrici di covarianza campionaria sono ortogonali e additive:

$$\mathbf{S}_T = \mathbf{S}_B + \mathbf{S}_W .$$

Questo approccio assume che le matrici di covarianza della popolazione,  $\Sigma_B$  e  $\Sigma_W$ , siano descritte da due modelli distinti, uno per la struttura di covarianza tra i gruppi e l'altro per la struttura di covarianza entro i gruppi.

Le strutture di covarianza nella popolazione, rispettivamente riferite ai modelli entro i gruppi e tra i gruppi, si possono specificare nel modo seguente:

$$\Sigma_W = \Lambda_W \Psi_W \Lambda_W' + \Theta_W ,$$

$$\Sigma_B = \Lambda_B \Psi_B \Lambda_B' + \Theta_B ,$$

dove  $\Psi_W$  e  $\Psi_B$  sono le matrici di covarianza fattoriali mentre  $\Theta_W$  e  $\Theta_B$  sono le matrici diagonali delle varianze dei residui, ciascuna riferita rispettivamente alle componenti entro e tra i gruppi. Anche i residui possono essere scomposti in una componente di gruppo ed una componente di varianza entro i gruppi:

$$\epsilon_{ij} = \epsilon_{Bj} + \epsilon_{Wij} ,$$

dove  $\epsilon_{Bj}$  è la componente casuale che cattura l'effetto di gruppo ed assume valore atteso nullo e  $\epsilon_{Wij}$  è la componente casuale che cattura l'effetto per le unità interne a ciascun gruppo. Le due componenti sono tra loro incorrelate,  $\text{cov}(\epsilon_{Bj}, \epsilon_{Wij}) = 0$ . Pertanto, anche la variazione residuale,  $\epsilon_{ij}$ , può essere scomposta in una componente di varianza di gruppo ed una componente di varianza entro i gruppi:

$$Var(\boldsymbol{\varepsilon}_{ij}) = \boldsymbol{\Theta}_B + \boldsymbol{\Theta}_W.$$

### 2.8.2 Il modello di misura a due livelli

La formulazione a due livelli del modello di misura riferito al vettore di variabili risposta latenti continue  $\mathbf{y}_{ij}^*$  rispettivamente per le differenze entro i gruppi e tra i gruppi si può esprimere, in maniera sintetica, nel modo seguente:

$$\begin{aligned}\mathbf{y}_{ij}^* &\sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_W(\boldsymbol{\theta}_W)), \\ \boldsymbol{\mu}_j &\sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_B(\boldsymbol{\theta}_B)).\end{aligned}$$

dove  $\boldsymbol{\mu}_j$  sono le intercette specifiche di ciascun gruppo mentre  $\boldsymbol{\mu}$  è l'intercetta generale. In alternativa, il modello di misura a due livelli si può esprimere in maniera più esplicita, utilizzando sempre una formulazione a due stadi:

$$\begin{aligned}\mathbf{y}_{ij}^* &= \boldsymbol{\mu}_j + \boldsymbol{\Lambda}_W \boldsymbol{\eta}_{Wij} + \boldsymbol{\varepsilon}_{Wij}, \\ \boldsymbol{\mu}_j &= \boldsymbol{\mu} + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{Bj} + \boldsymbol{\varepsilon}_{Bj}.\end{aligned}$$

La prima equazione rappresenta il modello di misura riferito al vettore di variabili risposta  $\mathbf{y}_{ij}^*$  per le differenze dovute principalmente all'effetto dei fattori individuali riassunti in  $\boldsymbol{\eta}_{Wij}$  ed include le intercette casuali  $\boldsymbol{\mu}_j$  che variano tra i diversi gruppi. La seconda equazione rappresenta il modello di misura per le intercette casuali  $\boldsymbol{\mu}_j$ , per le differenze dovute principalmente all'effetto dei fattori di gruppo riassunti in  $\boldsymbol{\eta}_{Bj}$ .

Il modello di misura si può esprimere anche in forma ridotta sostituendo l'equazione per le intercette casuali nell'equazione del modello di misura entro i gruppi:

$$\mathbf{y}_{ij}^* = \underbrace{\boldsymbol{\mu} + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{Bj} + \boldsymbol{\varepsilon}_{Bj}}_{\boldsymbol{\mu}_j} + \boldsymbol{\Lambda}_W \boldsymbol{\eta}_{Wij} + \boldsymbol{\varepsilon}_{Wij},$$

che corrisponde all'espressione riportata in Muthén e Satorra (1989) e in Muthén

(1994). Se si assume che il modello abbia un numero di fattori e matrici dei coefficienti fattoriali uguali ad entrambi i livelli, cioè  $\Lambda_{Bj} = \Lambda_{Wij} = \Lambda$ <sup>93</sup>, e che non ci siano residui tra i gruppi,  $\varepsilon_{Bj} = \mathbf{0}$ <sup>94</sup>, l'equazione precedente si semplifica ed assume la forma seguente:

$$\mathbf{y}_{ij}^* = \boldsymbol{\mu} + \Lambda(\boldsymbol{\eta}_{Bj} + \boldsymbol{\eta}_{Wij}) + \boldsymbol{\varepsilon}_{Wij}.$$

Questo modello semplificato, definito da Skrondal e Rabe-Hesketh (2004) come modello a componenti di varianza, assume una interpretazione interessante. Per questo modello, infatti,  $\boldsymbol{\eta}_{Bj}$  e  $\boldsymbol{\eta}_{Wij}$  rappresentano le componenti entro e tra i gruppi relativamente ai medesimi fattori (Bollen, Bauer, Christ e Edwards, 2010; Skrondal e Rabe-Hesketh, 2004). La variabile latente può essere scomposta come segue:

$$\eta_{ij} = \alpha + \eta_{Bj} + \eta_{Wij},$$

dove  $\alpha$  è il valore atteso di  $\eta_{ij}$ ,  $\eta_{Bj}$  è la componente casuale che cattura l'effetto di gruppo ed ha valore atteso uguale a zero mentre  $\eta_{Wij}$  è la componente casuale che varia tra le unità componenti il medesimo gruppo; le due componenti sono tra loro incorrelate,  $\text{cov}(\eta_{Bj}, \eta_{Wij}) = 0$ . Per catturare le differenze tra i gruppi sono quindi necessari solo i due parametri  $\alpha$  e la varianza di  $\eta_{Bj}$ . La varianza totale delle variabili latenti si può quindi suddividere in una componente entro i gruppi e una componente tra i gruppi. Questa scomposizione consente di determinare la quota di variabilità di un fattore dovuta alle differenze tra i gruppi rispetto alla variabilità dovuta alle differenze interne ai gruppi. In particolare, riferendosi ad un fattore generico, Muthén (1994) scompone la varianza totale nel modo seguente:

---

<sup>93</sup> Tale restrizione è appropriata solo se le strutture fattoriali dei modelli entro e tra i gruppi sono equivalenti ed implica che l'interpretazione delle variabili latenti è la stessa per entrambi i livelli.

<sup>94</sup> Tale assunzione è giustificata dal fatto che la variabilità oltre quella riassunta nei fattori di gruppo  $\boldsymbol{\eta}_{Bj}$  si può considerare come errore di misura e, come tale, non produce errore sistematico tra i gruppi (Bollen, Bauer, Christ e Edwards, 2010).

$$Var(\eta_{ij}) = \psi_T = \psi_B + \psi_W,$$

dove  $Var(\eta_{Bj}) = \psi_B$  e  $Var(\eta_{Wij}) = \psi_W$ . Da un punto di vista sostanziale, risulta interessante stimare la dimensione relativa della varianza tra i gruppi,  $\psi_B$ , rispetto alla varianza complessiva,  $\psi_T$ . Infatti, la covarianza fattoriale tra due unità,  $i$  e  $i'$ , appartenenti al medesimo gruppo, si può scrivere nel modo seguente:

$$Cov(\eta_{ij}, \eta_{ij}') = \underbrace{Cov(\eta_{Bj}, \eta_{Bj}')}_{\psi_B} + \underbrace{Cov(\eta_{Wij}, \eta_{Wij}')}_{0}.$$

Questo significa che i valori osservati per ciascuna unità appartenente al medesimo gruppo non sono indipendenti e che la grandezza di  $\psi_B$  misura il grado di non indipendenza. Come discusso da Muthén (1991), la controparte della correlazione intraclasse per le variabili latenti è data dal rapporto seguente:

$$ICC = \frac{\psi_B}{\psi_B + \psi_W}.$$

### 2.8.3 La stima pseudo bilanciata

La stima pseudo bilanciata è stata proposta da Muthén (1989) e da Muthén e Satorra (1989) ed è diversa a seconda che i gruppi siano bilanciati, aventi cioè la medesima numerosità, o non bilanciati.

Nel caso particolare di gruppi bilanciati, la stima è piuttosto semplice. Se i gruppi sono  $J$ , ciascuno con numerosità pari ad  $n$  e una dimensione campionaria complessiva  $N = nJ$ , si possono definire due matrici di covarianza campionarie: la matrice di covarianza *pooled-within*,  $\mathbf{S}_{PW}$ , e la matrice di covarianza *between* scalata,  $\mathbf{S}_B^*$ . La matrice di covarianza *pooled-within*,  $\mathbf{S}_{PW}$ , è lo stimatore ML della matrice di covarianza  $\Sigma_W$  (Muthén, 1989, 1990) e si ottiene nel modo seguente:

$$\mathbf{S}_{PW} = \frac{1}{N-J} \sum_{j=1}^J \sum_{i=1}^{n_j} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_j)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_j)' .$$

Poiché la matrice di covarianza *pooled-within*  $\mathbf{S}_{PW}$  è uno stimatore corretto della matrice di covarianza entro i gruppi della popolazione  $\Sigma_W$ , questa si può essere stimare direttamente con  $\mathbf{S}_{PW}$ . La matrice di covarianza tra i gruppi scalata campionaria,  $\mathbf{S}_B^*$ , è data da:

$$\mathbf{S}_B^* = \frac{1}{J-1} \sum_{j=1}^J n_j (\bar{\mathbf{y}}_j - \bar{\mathbf{y}})(\bar{\mathbf{y}}_j - \bar{\mathbf{y}})' ,$$

dove  $N = n_1 + n_2 + \dots + n_J$ . Muthén (1989, 1990) dimostra che  $E(\mathbf{S}_{PW}) = \Sigma_W$  e  $E(\mathbf{S}_B) = \Sigma_W + c\Sigma_B$ , dove  $c$  è la numerosità comune dei gruppi. Quindi,  $\mathbf{S}_{PW} = \hat{\Sigma}_W$  e  $\mathbf{S}_B^* = \hat{\Sigma}_W + c\hat{\Sigma}_B$ .

Nel caso più generale di gruppi non bilanciati, aventi cioè numerosità diversa,  $\mathbf{S}_{PW}$  è ancora lo stimatore ML di  $\Sigma_W$  mentre  $\mathbf{S}_B^*$  adesso stima una matrice per ciascun insieme di gruppi aventi numerosità diversa  $d$ :

$$\mathbf{S}_{Bd}^* = \hat{\Sigma}_W + c_d \hat{\Sigma}_B ,$$

dove l'espressione vale per ciascun insieme distinto di gruppi aventi numerosità uguale ad  $n_d$  e  $c_d = n_d$  (Muthén, 1990, 1994). La stima di massima verosimiglianza ad informazione completa (FIML) in presenza di gruppi non bilanciati implica quindi la specificazione di un modello di covarianza tra i gruppi separato per ciascun gruppo avente numerosità diversa, ognuno con un diverso parametro di scala  $c_d$ . Poiché la stima di massima verosimiglianza risulta problematica, Muthén (1989, 1990) propose di ignorare la diversa numerosità dei gruppi e di calcolare un'unica matrice  $\mathbf{S}_B^*$ , considerando la numerosità campionaria media, ottenuta nel modo seguente:

$$c^* = \frac{1}{N(G-1)} \left( N^2 - \sum_{j=1}^G n_j^2 \right).$$

Il risultato è uno stimatore di massima verosimiglianza ad informazioni limitata, chiamato soluzione pseudo bilanciata (McDonald, 1994) o MUML<sup>95</sup> (Muthén, 1989, 1994).

Muthén (1989, 1994) dimostra che  $\mathbf{S}_B^*$  è uno stimatore consistente e non distorto di  $\Sigma_W + c\Sigma_B$ , cioè  $E(\mathbf{S}_B) = \Sigma_W + c\Sigma_B$ . Questo significa che per grandi campioni (sia delle unità statistiche che dei gruppi),  $\mathbf{S}_B^*$  è una stima prossima a  $\Sigma_B$  e la soluzione pseudo bilanciata produce una buona approssimazione.

Lo stimatore ML proposto da Muthén (1989, 1990)  $\hat{\boldsymbol{\theta}}$  minimizza la seguente funzione di adattamento:

$$F_{MUML}(\boldsymbol{\theta}) = Gtr\left(\left((\Sigma_W + c\Sigma_B)^{-1}\mathbf{S}_B\right) - \log\left|(\Sigma_W + c\Sigma_B)^{-1}\mathbf{S}_B\right| - p\right) + (N-j)\left(tr\left(\Sigma_W^{-1}\mathbf{S}_W\right) - \log\left|\Sigma_W^{-1}\mathbf{S}_W\right| - p\right),$$

dove  $p$  è il numero di variabili risposta osservate.

Studi di simulazione (Hox e Maas, 2001; Hox, Maas e Brinkhuis, 2010) hanno dimostrato che la stima della parte entro i gruppi del modello non pone problemi in nessuna condizione simulata. Per quanto riguarda la parte tra i gruppi, la stima dei coefficienti fattoriali è generalmente accurata. Invece, le varianze residue sono sottostimate mentre gli errori standard sono generalmente troppo piccoli. Inoltre, Yuan e Hayashi (2005) hanno dimostrato analiticamente che gli errori standard e la statistica chi-quadrato conducono a inferenze corrette solo nel caso in cui la numerosità campionaria tende all'infinito ed il coefficiente di variazione della dimensione dei gruppi tende a zero. Sia gli studi di simulazione che i lavori analitici sono d'accordo nel ritenere che numerosità campionarie più grandi non migliorano l'accuratezza nel caso di dati non bilanciati. Così, quando i dati sono decisamente non bilanciati, lo stimatore MUML produce errori standard e test di significatività distorti e la distorsione non si riduce all'aumentare della dimensione campionaria.

---

<sup>95</sup> MUML è l'acronimo di *Muthén's ML*.

#### *2.8.4 La stima di massima verosimiglianza ad informazione completa*

Lo stimatore MUML prevede la stima di due matrici di covarianza separate (entro e tra i gruppi) e consente di stimare solo modelli ad equazioni strutturali ad intercetta casuale. Dopo lo stimatore MUML, sono stati proposti diversi metodi di stima basati sul metodo di massima verosimiglianza ad informazione completa (Lee e Poon, 1998; Bentler e Liang, 2003; Asparouhov e Muthén, 2003; Liang e Bentler, 2004; du Toit e du Toit, 2004). Alcuni di questi stimatori si basano sull'algoritmo EM (Dempster, Laird e Rubin, 1977) per la stima in presenza di dati mancanti. Nel contesto dei modelli multilivello a variabili latenti (MLLVM), l'algoritmo EM è stato utilizzato da Lee e Poon (1998), Bentler e Liang (2003) e Liang e Bentler (2004) per i modelli ad equazioni strutturali a due livelli dove la parte between di una variabile è considerata come mancante. Asparouhov e Muthén (2003) hanno sviluppato tre stimatori di massima verosimiglianza basati sull'algoritmo EM che combina gli approcci di Lee e Poon e Bentler e Liang mentre du Toit e du Toit (2004) hanno utilizzato l'algoritmo Fisher scoring.

La teoria di base della stima ML per l'analisi dei modelli ad equazioni strutturali multilivello in condizioni di normalità multivariata si trova in Lee (1990) che ha proposto un modello ad equazioni strutturali con disegno campionario non bilanciato utilizzando i metodi di stima di massima verosimiglianza e dei minimi quadrati generalizzati, illustrando la relazione tra questi. Tale modello, a differenza di quelli proposti da McDonald e Goldstein (1989) e da Muthén e Satorra (1989), non prevede la possibilità di considerare variabili a livello di gruppo. Dal punto di vista computazionale, le stime sono ottenute con gli algoritmi scoring e Gauss-Newton.

Lee e Poon (1998) hanno stimato un modello ad equazioni strutturali a due livelli con disegno non bilanciato trattando i vettori casuali delle variabili latenti di secondo livello come dati mancanti ipotetici. La soluzione ML è ottenuta applicando l'algoritmo EM la cui performance può essere migliorata utilizzando procedure accelerate come l'algoritmo EM gradiente e l'algoritmo EM Quasi-Newton per ottenere una convergenza più rapida.

Asparouhov e Muthén (2003) hanno sviluppato tre stimatori ML basati sull'algoritmo EM che combinano gli approcci di Lee e Poon (1998) e Bentler e Liang, (2003). Questi stimatori si distinguono per l'approccio utilizzato nel calcolo degli errori

standard e sono definiti con gli acronimi MLF, ML e MLR (vedi paragrafo 2.4.3). In particolare, lo stimatore MLR produce le medesime stime dei parametri del metodo ML mentre gli errori standard sono stimatori sandwich del tipo Huber-White (Huber, 1967; White, 1982) che utilizzano le varianze residue osservate per correggere gli errori standard asintotici. I test di adattamento chi-quadrato sono invece calcolati con le correzioni Satorra-Bentler (1994) e Yuan-Bentler (1998)<sup>96</sup>. Lo stimatore MLR è uno stimatore ML robusto ad informazione completa e produce una matrice di covarianza delle stime corretta che non dipende dall'ipotesi di normalità ed un test di adattamento chi-quadrato robusto. In presenza di dati aventi una struttura multilivello, gli errori standard ed i test di adattamento robusti sono di solito più accurati dei test asintotici e sono in grado di proteggere i risultati nei confronti di eterogeneità non modellata che può verificarsi per l'omissione di un livello di aggregazione o a causa di un modello mal specificato.

Liang e Bentler (2004), dopo aver discusso le similarità e le differenze tra le diverse formulazioni dei modelli ad equazioni strutturali a due livelli presenti in letteratura, hanno proposto un algoritmo EM efficiente dal punto di vista computazionale per la stima ML dei parametri in caso di disegno non bilanciato e casi mancanti a caso (MAR). Tale formulazione è una generalizzazione del modello studiato da Lee (1990) e da Lee e Poon (1998). L'algoritmo proposto è applicabile alla stima di massima verosimiglianza completa dei modelli ad equazioni strutturali a due livelli con o senza osservazioni a livello di gruppo. In particolare, il nuovo algoritmo EM può essere implementato per la stima dei modelli ad equazioni strutturali a due livelli per i modelli proposti in McDonald e Goldstein (1989), Muthén (1989; 1990, 1994, 1997), Muthén e Satorra (1989) McDonald (1994), Lee e Poon (1998) con matrici di covarianza entro i gruppi uguali e du Toit e du Toit (2004).

du Toit e du Toit (2004) hanno proposto un modello ad equazioni strutturali a due livelli con disegno non bilanciato e dati mancanti ad entrambi i livelli della gerarchia, simile a quello di Liang e Bentler (2004), in cui la differenza principale è costituita dalla procedura di stima. Tale approccio utilizza l'algoritmo Fisher scoring per ottenere le stime ML. Un vantaggio di questo metodo è che utilizza i valori attesi delle derivate seconde quindi gli errori standard delle stime dei parametri sono subito

---

<sup>96</sup> Tale correzione è impiegata dal software Mplus (Muthén e Muthén, 1998-2007).



disponibili. Tali autori propongono un algoritmo per la stima di massima verosimiglianza ad informazione completa (FIML) e derivano una funzione di discrepanza e test di adattamento basati sulla verosimiglianza.

Metha e Neale (2005) hanno utilizzato la funzione di verosimiglianza individuale (vedi paragrafo 2.4.3) per specificare e stimare modelli piuttosto complicati che includono anche strutture dati non bilanciate in cui i gruppi sono considerati come osservazioni e le unità all'interno dei gruppi come variabili. In questo modo i dati non bilanciati rappresentano gruppi con un numero di unità diverse e sono inclusi come dati incompleti in un modello ad equazioni strutturali standard. L'utilizzo delle funzioni di verosimiglianza individuali consente di trattare i dati mancanti senza dover ricorrere alla eliminazione "listwise" e facilita il calcolo del vettore delle medie e della matrice di covarianza implicata dal modello per ciascun insieme di risposte individuali<sup>97</sup>. Le variabili mancanti sono eliminate dal vettore delle medie e dalla matrice di covarianza individuale. In questo modo la dimensione della matrice di covarianza varia tra gli individui e dipende dal numero di osservazioni presenti per l'individuo considerato.

### *2.8.5 La stima con variabili osservate categoriche*

I modelli multilivello con variabili dipendenti osservate distribuite normalmente si possono stimare con il metodo di massima verosimiglianza ricorrendo all'algoritmo EM (Raudenbush e Bryk, 2002). Questo metodo di stima si può utilizzare senza particolari problemi anche quando gli effetti casuali sono piuttosto numerosi.

Quando le variabili dipendenti osservate non sono distribuite normalmente, i modelli multilivello si possono stimare con il metodo di massima verosimiglianza ricorrendo all'algoritmo EM ma gli effetti casuali devono essere integrati con una procedura numerica (Muthén e Asparouhov, 2006). Tuttavia, se gli effetti casuali sono numerosi, l'integrazione numerica diventa molto pesante dal punto di vista computazionale e la stima del modello richiede molto tempo. Nelle applicazioni pratiche, di solito questi modelli sono stimati con l'integrazione numerica utilizzando un numero minore di punti di integrazione per ogni dimensione o ricorrendo all'integrazione Monte Carlo. Tuttavia, questi metodi presentano tipicamente problemi

---

<sup>97</sup> Nell'ipotesi di dati mancanti a caso, condizionatamente ai dati osservati, le medie e le covarianze attese individuali per le variabili con osservazioni mancanti per un individuo sono uguali alle corrispondenti medie e covarianze per un individuo con dati completi (Metha e Neale, 2005).

di convergenza e mancanza di precisione delle stime ed i tempi di calcolo molto lunghi li rendono quasi inapplicabili. In alternativa, si possono utilizzare i metodi di stima bayesiana come il MCMC ma anche questi sono molto pesanti dal punto di vista computazionale ed hanno una prestazione simile al metodo di massima verosimiglianza con l'integrazione Monte Carlo.

Per l'analisi dei modelli ad equazioni strutturali a due livelli con variabili risposta osservate binarie, categoriche ordinali e continue, Lee e Shi (2001) hanno utilizzato un approccio di massima verosimiglianza considerando le variabili latenti come dati mancanti ipotetici ed utilizzando un algoritmo EM Monte Carlo (MCEM; Wei e Tanner, 1990). Questo modello è stato successivamente esteso da Lee e Song (2004) per analizzare relazioni non lineari di primo e di secondo livello tra le variabili latenti e gli effetti di variabili esplicative di primo e di secondo livello nel modello di misura e nel modello a variabili latenti<sup>98</sup>.

Asparouhov e Muthén (2007) hanno proposto invece un metodo di stima ad informazione limitata basato sui minimi quadrati ponderati che può essere utilizzato per stimare modelli ad equazioni strutturali a due livelli con variabili osservate binarie, categoriche ordinali, censurate, continue o combinazioni tra queste. Questo approccio può considerarsi una estensione del metodo di stima dei minimi quadrati ponderati (WLS) proposto da Muthén (1984) con il successivo contributo di Muthén, du Toit, e Spisic (1997). Lo stimatore dei minimi quadrati ponderati si può utilizzare per stimare modelli con un numero qualsiasi di effetti casuali senza che questo comporti un allungamento dei tempi di calcolo o comprometta la precisione delle stime. In pratica, il metodo proposto sostituisce un sistema di stima molto complesso basato su integrazioni multiple ad elevate dimensioni con una serie di modelli più semplici con integrazioni ad una o due dimensioni. Tale modello è sufficientemente generale perché consente di considerare gli effetti di variabili esplicative di primo e di secondo livello nel modello a variabili latenti ed è implementato nel software *Mplus* (versione 5 e successive).

---

<sup>98</sup> Per la stima di tale modello, gli autori hanno anche implementato un software in C disponibile su richiesta.

### 2.8.6 Il metodo dei minimi quadrati ponderati ad informazione limitata

Il punto di partenza è il vettore  $1 \times p$  di variabili risposta osservate  $\mathbf{y}'_{ij} = (y_{1ij}, y_{2ij}, \dots, y_{pij})$  riferite all'unità statistica di primo livello  $i$ , con  $i = 1, \dots, N_j$ , inclusa nell'unità di secondo livello (gruppo)  $j$ , con  $j = 1, \dots, C$ , dove  $N_j$  è il numero di unità appartenenti al gruppo  $j$ . Inoltre, si definisce il vettore  $1 \times q_1$  di variabili esplicative osservate  $\mathbf{x}'_{wij} = (x_{w1ij}, x_{w2ij}, \dots, x_{wq_{1ij}})$  riferite all'unità di primo livello  $i$  appartenente al gruppo  $j$ , ed il vettore  $1 \times q_2$  di variabili esplicative osservate  $\mathbf{x}'_{bj} = (x_{b1j}, x_{b2j}, \dots, x_{bq_{2j}})$ , riferito all'unità di secondo livello  $j$ , dove gli indici  $w$  e  $b$  stanno a significare rispettivamente “within” e “between”.

Il modello a due livelli considera variabili risposta osservate dicotomiche, categoriche ordinali, continue e censurate analogamente a quanto specificato da Muthén (1984) per i modelli ad equazioni strutturali ad un livello, definendo una variabile latente continua distribuita normalmente per ciascuna variabile osservata. In particolare, la variabile latente sottostante, indicata con  $y^*_{pij}$ , è definita in maniera diversa a seconda del tipo di variabile osservata  $y_{pij}$  a cui si riferisce. Se la variabile osservata è distribuita normalmente allora la variabile latente si osserva direttamente,  $y^*_{pij} = y_{pij}$ . Se la variabile osservata è categorica ordinale, la variabile latente è definita a meno di una serie di parametri soglia, indicati con  $\tau_{pk}$ :

$$y_{pij} = k \Leftrightarrow \tau_{pk} < y^*_{pij} < \tau_{pk+1},$$

dove  $k = 0, 1, \dots, K-1$  indica la modalità della variabile categorica osservata mentre i parametri soglia sono  $-\infty = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < \tau_K = +\infty$ <sup>99</sup>. Se la variabile osservata è censurata al punto  $c_p$ , la variabile latente si definisce nel modo seguente:

$$y_{pij} = \begin{cases} y^*_{pij} & \text{se } y^*_{pij} > c_p \\ c_p & \text{se } y^*_{pij} < c_p \end{cases}.$$

<sup>99</sup> La variabile risposta dicotomica si può considerare come caso particolare di una variabile categorica ordinale in cui  $C = 2$ .

La variabile latente  $\mathbf{y}_{ij}^*$  è scomposta in due variabili latenti indipendenti distribuite normalmente e additive (Muthén, 1994). La prima,  $\mathbf{y}_{bj}$ , rappresenta le differenze tra le unità di secondo livello (intercette casuali o effetti di gruppo), la seconda,  $\mathbf{y}_{wij}$ , rappresenta le differenze tra le unità di primo livello all'interno di ciascuna unità di secondo livello (effetto individuale):

$$\mathbf{y}_{ij}^* = \mathbf{y}_{wij} + \mathbf{y}_{bj}.$$

Il modello a due livelli si ottiene specificando due modelli ad equazioni strutturali separati, rispettivamente per  $\mathbf{y}_{bpj}$  e  $\mathbf{y}_{wpj}$ . Per questi, si definisce rispettivamente il vettore  $1 \times m_1$  di variabili latenti per le unità di primo livello  $\boldsymbol{\eta}'_{wij} = (\eta_{1ij}, \eta_{2ij}, \dots, \eta_{m_{1ij}})$  ed il vettore  $1 \times m_2$  di variabili latenti per le unità di secondo livello  $\boldsymbol{\eta}'_{bj} = (\eta_{1bj}, \eta_{2bj}, \dots, \eta_{m_{2bj}})$ . Le variabili latenti di primo e di secondo livello sono distribuite normalmente.

Per le unità di primo livello, il modello di misura ed il modello a variabili latenti sono specificati rispettivamente nel modo seguente:

$$\begin{aligned}\mathbf{y}_{wpj} &= \mathbf{\Lambda}_w \boldsymbol{\eta}_{wij} + \boldsymbol{\varepsilon}_{wij}, \\ \boldsymbol{\eta}_{wij} &= \mathbf{B}_w \boldsymbol{\eta}_{wij} + \mathbf{\Gamma}_w \mathbf{x}_{wij} + \boldsymbol{\zeta}_{wij},\end{aligned}$$

dove  $\mathbf{\Lambda}_w$  è la matrice dei coefficienti fattoriali,  $\boldsymbol{\varepsilon}_{wij}$  è il vettore casuale dei residui,  $\mathbf{B}_w$  la matrice dei coefficienti riferiti alle variabili latenti,  $\mathbf{\Gamma}_w$  la matrice dei coefficienti tra le variabili latenti e le variabili esplicative osservate mentre  $\boldsymbol{\zeta}_{wij}$  è il vettore dei residui.

Per le unità di secondo livello, il modello di misura ed il modello a variabili latenti sono invece specificati rispettivamente nel modo seguente:

$$\mathbf{y}_{bj} = \mathbf{v}_b + \mathbf{\Lambda}_b \boldsymbol{\eta}_{bj} + \boldsymbol{\varepsilon}_{bj}$$

$$\boldsymbol{\eta}_{bj} = \boldsymbol{\alpha}_b + \mathbf{B}_b \boldsymbol{\eta}_{bj} + \boldsymbol{\Gamma}_b x_{bj} + \boldsymbol{\zeta}_{bj}.$$

I parametri da stimare sono compresi nelle matrici  $\boldsymbol{\Lambda}_w$ ,  $\boldsymbol{\Gamma}_w$ ,  $\mathbf{B}_w$ ,  $\boldsymbol{\Lambda}_b$ ,  $\boldsymbol{\Gamma}_b$  e  $\mathbf{B}_b$ , e nei vettori  $\mathbf{v}_b$ ,  $\boldsymbol{\alpha}_b$  e  $\boldsymbol{\tau}_{pk}$ . I residui di primo e secondo livello sono distribuiti normalmente con media nulla e una matrice di covarianza specificata,  $\boldsymbol{\varepsilon}_{wij} \sim N(\mathbf{0}, \boldsymbol{\Theta}_w)$ ,  $\boldsymbol{\zeta}_{wij} \sim N(\mathbf{0}, \boldsymbol{\Psi}_w)$ ,  $\boldsymbol{\varepsilon}_{bj} \sim N(\mathbf{0}, \boldsymbol{\Theta}_b)$  e  $\boldsymbol{\zeta}_{bj} \sim N(\mathbf{0}, \boldsymbol{\Psi}_b)$ . Per problemi di identificazione del modello, la varianza dei residui di primo livello  $\boldsymbol{\varepsilon}_{wij}$  è fissata al valore uno per le variabili categoriche. Il modello strutturale può essere scritto, in termini sintetici, nel modo seguente:

$$\mathbf{y}_{wij} = \boldsymbol{\Pi}_w x_{wij} + \boldsymbol{\omega}_{wij},$$

$$\mathbf{y}_{bj} = \boldsymbol{\mu}_b + \boldsymbol{\Pi}_b x_{bj} + \boldsymbol{\omega}_{bj},$$

dove  $\boldsymbol{\Pi}_w = \boldsymbol{\Lambda}_w (\mathbf{I} - \mathbf{B}_w)^{-1} \boldsymbol{\Gamma}_w$ ,  $\boldsymbol{\Pi}_b = \boldsymbol{\Lambda}_b (\mathbf{I} - \mathbf{B}_b)^{-1} \boldsymbol{\Gamma}_b$  mentre  $\boldsymbol{\mu}_b = \mathbf{v}_b + \boldsymbol{\Lambda}_b (\mathbf{I} - \mathbf{B}_b)^{-1} \boldsymbol{\alpha}_b$ . I residui sono distribuiti normalmente  $\boldsymbol{\omega}_{bj} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_b)$  e  $\boldsymbol{\omega}_{wij} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_w)$ . Per le variabili categoriche la varianza di  $\omega_{wpj}$  è fissata ad uno mentre il parametro  $\mu_{pb}$  è fissato a zero per problemi di identificazione del modello. Le strutture di covarianza rispettivamente per il modello entro e tra i gruppi assumono la forma seguente:

$$\boldsymbol{\Sigma}_w = \boldsymbol{\Lambda}_w (\mathbf{I} - \mathbf{B}_w)^{-1} \boldsymbol{\Psi}_w (\mathbf{I} - \mathbf{B}_w)^{-1'} \boldsymbol{\Lambda}_w' + \boldsymbol{\Theta}_w,$$

$$\boldsymbol{\Sigma}_b = \boldsymbol{\Lambda}_b (\mathbf{I} - \mathbf{B}_b)^{-1} \boldsymbol{\Psi}_b (\mathbf{I} - \mathbf{B}_b)^{-1'} \boldsymbol{\Lambda}_b' + \boldsymbol{\Theta}_b.$$

La stima di questo modello viene effettuata con una procedura ad informazione limitata a due stadi (paragrafo 2.4.6). Nel primo stadio si stimano i  $p$  modelli univariati utilizzando il metodo di massima verosimiglianza descritto in Muthén e Asparouhov (2006) e si ottengono i parametri riassunti in  $\boldsymbol{\Pi}_w$ ,  $\boldsymbol{\Pi}_b$ ,  $\boldsymbol{\mu}_b$ ,  $\boldsymbol{\tau}_{pk}$ ,  $\boldsymbol{\Sigma}_w$  e  $\boldsymbol{\Sigma}_b$ . Nel secondo stadio si stimano invece tutte le possibili coppie di modelli bivariati, condizionatamente ai valori dei parametri ottenuti nella prima fase. In questo modo, per il modello

bivariato tra le due variabili  $p_1$  e  $p_2$  è necessario stimare soltanto due parametri,  $\Sigma_{wp_1p_2}$  e  $\Sigma_{bp_1p_2}$ . Questo secondo stadio della procedura viene realizzato con il metodo di massima verosimiglianza multilivello descritto in Muthén e Asparouhov (2006), agevolando la massimizzazione della verosimiglianza bivariata descritta da Olsson (1979) per il calcolo dei coefficienti di correlazione policorica e da Olsson, Dragsow e Dorans (1982) per il calcolo del coefficiente di correlazione poliseriale.

La procedura di stima univariata e bivariata descritta richiede un'integrazione numerica per ogni variabile non distribuita normalmente e segue il metodo descritto da Raudenbush e Bryk (2002) per le variabili aventi una distribuzione normale. In particolare, la stima univariata di una variabile categorica utilizza una dimensione di integrazione numerica, la stima bivariata di due variabili categoriche utilizza due dimensioni di integrazione numerica mentre la stima bivariata di una variabile distribuita normalmente ed una variabile categorica utilizza una dimensione di integrazione numerica.

La matrice di covarianza asintotica dei parametri ottenuti nei primi due stadi della procedura viene calcolata come descritto da Muthén (1984) e da Muthén e Satorra (1995) per i modelli ad equazioni strutturali ad un livello (paragrafo 2.4.6). Tale matrice si indica con  $\mathbf{G}$  e comprende gli elementi del vettore delle derivate prime corrispondenti ai primi due stadi della procedura. I parametri da stimare sono riassunti per comodità nel vettore  $\mathbf{s}$ .

L'ultimo stadio della procedura di stima è la minimizzazione della funzione di adattamento rispetto ai parametri del modello. La funzione di adattamento è quella dei minimi quadrati ponderati, dove  $\mathbf{W}$  è la matrice dei pesi avente la stessa dimensione del vettore  $\mathbf{s}$ :

$$F_{WLS} = (\mathbf{s} - \boldsymbol{\sigma})\mathbf{W}(\mathbf{s} - \boldsymbol{\sigma})'.$$

Utilizzando una diversa matrice dei pesi  $\mathbf{W}$  si ottengono stimatori diversi. Se la matrice dei pesi è la matrice identità,  $\mathbf{W} = \mathbf{I}$ , si ottiene lo stimatore dei minimi quadrati non ponderati (ULS). Se la matrice dei pesi è l'inversa della matrice di covarianza asintotica dei parametri stimati nei primi due stadi della procedura,  $\mathbf{W} = \mathbf{G}^{-1}$ , si ottiene invece lo

stimatore dei minimi quadrati ponderati (WLS).

Tuttavia, spesso il modello non ristretto presenta un numero elevato di parametri e la dimensione della matrice  $\mathbf{G}$  è più grande del numero di gruppi (unità di secondo livello) nel campione. Questo comporta una matrice  $\mathbf{G}$  singolare ed uno stimatore WLS indefinito. In questo caso, Muthén, du Toit e Spisic (1997) hanno suggerito di considerare una matrice dei pesi diagonale,  $\mathbf{W} = \mathbf{G}_0^{-1}$ , dove  $\mathbf{G}_0$  è composta dai soli elementi sulla diagonale principale della matrice di covarianza asintotica  $\mathbf{G}$ . Tale stimatore si definisce stimatore dei minimi quadrati diagonali, indicato in Mplus con l'acronimo WLSM o WLSMV, dove la differenza tra l'uno e l'altro risiede unicamente nel metodo di calcolo della statistica chi-quadrato (Muthén, du Toit e Spisic, 1997) (vedi paragrafo 2.6.6).

Asparouhov e Muthén (2007) hanno realizzato uno studio di simulazione per analizzare la prestazione dello stimatore dei minimi quadrati ponderati (WLSM) e confrontarla con quella dello stimatore di massima verosimiglianza (ML). Lo studio è stato effettuato considerando prima un modello fattoriale confermativo a due livelli con variabili risposta categoriche a cinque modalità ordinate e poi lo stesso modello con variabili risposta continue. Il tasso di convergenza dello stimatore WLSM è risultato pari al 100% sia nel modello con variabili risposta categoriche sia in quello con variabili risposta continue mentre il tasso di convergenza dello stimatore ML è risultato del 100% solo nel modello con variabili risposta continue e appena del 47% in quello con variabili risposta categoriche. Per quanto riguarda l'efficienza delle stime, dal punto di vista teorico lo stimatore ML è il più efficiente ma dal punto di vista pratico, a causa del ricorso all'integrazione numerica, le sue stime sono approssimate. In conclusione, quando le variabili risposta osservate sono categoriche, i risultati ottenuti mostrano che le stime ottenute con lo stimatore WLSM sono più efficienti e meno distorte di quelle ottenute con lo stimatore ML. Invece, quando le variabili risposta osservate sono continue, entrambi gli stimatori non sono distorti.

## 2.9 Modelli ad equazioni strutturali multilivello generalizzati

L'analisi tradizionale dei modelli ad equazioni strutturali a due livelli basata sulla formulazione di modelli separati per le matrici di covarianza entro i gruppi e tra i gruppi presenta alcune limitazioni: i) non permette di specificare relazioni tra variabili (latenti ed osservate) misurate a livelli diversi; ii) non permette di considerare più di due livelli di analisi; iii) le variabili risposta osservate (indicatori) non possono variare a livelli diversi; iv) non si possono prevedere coefficienti casuali per le variabili di secondo livello; v) consente l'analisi di variabili risposta continue o categoriche per le quali si ipotizza l'esistenza di una variabili latente continua sottostante.

Per superare queste limitazioni, Rabe-Hesketh, Skrondal e Pickles (2004) hanno proposto i modelli ad equazioni strutturali generalizzati multilivello, un approccio unitario che consente di specificare modelli ad equazioni strutturali multilivello con variabili latenti senza ricorrere alla stima di matrici di covarianza separate. Tale approccio rientra nell'ambito dei modelli lineari generalizzati misti a variabili latenti (GLLAMM)<sup>100</sup> e consente di combinare le caratteristiche dei modelli lineari generalizzati misti con quelle dei modelli ad equazioni strutturali a variabili latenti.

I modelli lineari generalizzati misti con variabili latenti sono composti da un modello di risposta (o modello di misura) (paragrafo 2.9.1) e da un modello a variabili latenti (paragrafo 2.9.2). Il modello di risposta rappresenta una estensione dei modelli lineari generalizzati misti, permette di modellare un'ampia gamma di variabili risposta e di incorporare strutture fattoriali in aggiunta ad intercette e coefficienti casuali. Il modello a variabili latenti è simile a quello dei modelli ad equazioni strutturali tradizionali con la differenza che consente di includere variabili latenti e variabili osservate che variano a livelli differenti.

I modelli ad equazioni strutturali generalizzati multilivello sono in grado di trattare; i) un numero arbitrario di livelli; ii) dati mancanti mancanti a caso (MAR) e non a caso (NMAR); iii) disegni multilivello non bilanciati; iv) coefficienti casuali con variabili esplicative non bilanciate; v) strutture fattoriali generali (con coefficienti

---

<sup>100</sup> L'acronimo GLLAMM significa *Generalized Linear Latent And Mixed Models*. Lo sviluppo dell'ambiente GLLAMM è avvenuto in parallelo a quello del programma statistico "gllamm" che lavora in Stata, ed è disponibile per il download al seguente indirizzo internet: [www.gllamm.org](http://www.gllamm.org). Tale programma può stimare tutti i modelli previsti dall'ambiente GLLAMM ad eccezione, per il momento, dei modelli che prevedono variabili latenti discrete e continue. Inoltre, per le variabili latenti continue l'unica distribuzione disponibile è la distribuzione normale multivariata.



fattoriali incogniti); vi) regressioni tra le variabili latenti che variano a livelli differenti; vii) variabili risposta osservate continue, frequenze, risposte categoriche ordinali e non ordinali, graduatorie, durate (a tempo continuo o discreto) e combinazioni tra queste. Tale approccio risulta quindi molto flessibile ed in grado di accomodare strutture fattoriali piuttosto complesse.

### 2.9.1 Il modello di risposta

Il modello di risposta (o di misura) è un modello lineare generalizzato misto in cui il valore atteso della variabile risposta,  $E(y_{ij})$ , condizionato alle variabili latenti, riassunte in  $\boldsymbol{\eta}_j$ , e alle variabili esplicative,  $\mathbf{x}_j$  e  $\mathbf{z}_j$ , ed è specificato da un predittore lineare,  $\nu_{ij}$ , una funzione legame,  $g(\cdot)$ , e dalla distribuzione della variabile risposta nell'ambito della famiglia esponenziale:

$$g(E[y_{ij} | \mathbf{x}_j, \mathbf{z}_j, \boldsymbol{\eta}_j]) = \nu_{ij},$$

dove la funzione legame può essere una qualunque tra quelle previste per i modelli lineari generalizzati misti.

Nel caso di un modello a  $L$  livelli e  $M_l$  variabili latenti a livello  $l > 1$ , il predittore lineare assume la forma seguente:

$$\nu = \mathbf{B}'\mathbf{x} + \sum_{l=2}^L \sum_{m=1}^{M_l} \boldsymbol{\eta}_m^{(l)} \mathbf{z}_m^{(l)} \boldsymbol{\lambda}_m^{(l)},$$

dove sono stati omessi per semplicità gli indici (pedici) per le unità ai diversi livelli ed il primo elemento di  $\boldsymbol{\lambda}_m^{(l)}$  è tipicamente fissato ad assumere valore 1, cioè  $\lambda_m^{(l)} = 1$  per problemi di identificazione. Gli elementi del vettore  $\mathbf{x}$  sono le variabili esplicative associate agli effetti fissi,  $\mathbf{B}$ . La variabile latente  $m$ -esima al livello  $l$ ,  $\boldsymbol{\eta}_m^{(l)}$ , è moltiplicata per una combinazione lineare di variabili esplicative,  $\mathbf{z}_m^{(l)} \boldsymbol{\lambda}_m^{(l)}$ , dove  $\boldsymbol{\lambda}_m^{(l)}$  sono i parametri (o coefficienti fattoriali).

Il vettore di variabili latenti a livello  $l$  per una data unità a livello  $l$  è indicato con  $\boldsymbol{\eta}^{(l)} = (\boldsymbol{\eta}_1^{(l)}, \dots, \boldsymbol{\eta}_{M_l}^{(l)})'$  ed il corrispondente vettore di variabili esplicative è  $\mathbf{z}_m^{(l)} = (\mathbf{z}_1^{(l)}, \dots, \mathbf{z}_{M_l}^{(l)})'$ . Inoltre, si definisce il vettore di variabili latenti a tutti i livelli per una unità a livello 2,  $\boldsymbol{\eta} = (\boldsymbol{\eta}^{(2)}, \dots, \boldsymbol{\eta}^{(L)})'$  con il vettore di variabili esplicative corrispondente,  $\mathbf{z} = (\mathbf{z}^{(2)}, \dots, \mathbf{z}^{(L)})'$ . Non sono previste, invece, variabili latenti a livello 1, essendo il livello riservato all'errore della distribuzione condizionale scelta.

### 2.9.2 La tipologia delle variabili risposta

Nel caso di risposte continue, si assume di solito un funzione legame identità e una distribuzione normale:

$$y = \nu + \varepsilon,$$

con  $f(\varepsilon) = (2\pi\sigma^2)^{-1/2} \exp(-\varepsilon^2/(2\sigma^2))$ . La funzione di densità condizionata diventa:  $f(y | \mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) = \sigma^{-1} \phi(\nu\sigma^{-1})$ , dove  $\phi$  indica la densità normale standard. L'eteroschedasticità può, invece, essere modellata nel modo seguente,  $\log(\sigma) = \boldsymbol{\alpha}'\mathbf{z}^{(1)}$ , dove  $\sigma$  è lo scarto quadratico medio di  $\varepsilon$ ,  $\boldsymbol{\alpha}$  i parametri e  $\mathbf{z}^{(1)}$  le covariate.

Nel caso di risposte ordinali, dicotomiche e durate a tempo discreto, il valore atteso della risposta si modifica in modo tale che la funzione legame si possa applicare alle probabilità cumulate invece che al valore atteso:

$$g(P[y \leq y_s | \mathbf{x}, \mathbf{z}, \boldsymbol{\eta}]) = k_s - \nu, \quad s = 1, \dots, S-1$$

dove  $y_s$ ,  $s = 1, 2, \dots, S$  sono le modalità di risposta e  $k_s$ ,  $s = 2, \dots, S-1$  con  $0 = k_1 < k_2 < \dots < k_{S-1}$  sono i parametri soglia da stimare. Come funzione legame si può assumere indifferentemente una distribuzione multinomiale, logit, probit o log-log complementare. In alternativa, i modelli si possono definire utilizzando il concetto di variabile risposta latente continua:

$$y^* = \nu + \varepsilon^*,$$

con risposte ordinali o dicotomiche generate tramite un modello soglia:

$$y = y_s \quad \text{se} \quad k_{s-1} < y^* \leq k_s, \quad k_0 = -\infty, \quad k_1 = 0, \quad k_S = \infty.$$

Le funzione legame logit, probit e log-log complementare corrispondono a specificare rispettivamente  $f(\varepsilon) = \exp(-\varepsilon)[1 + \exp(-\varepsilon)]^{-2}$ ,  $f(\varepsilon) = (2\pi)^{-1/2} \exp(-1/2 \varepsilon^2)$  e  $f(\varepsilon) = (\varepsilon - \exp(\varepsilon))$ . Se la varianza di  $\varepsilon$  è identificabile, si possono utilizzare versioni scalate di queste funzioni di densità che corrispondono a funzioni legame scalate. La distribuzione condizionata diventa:

$$f(y = y_s | \mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) = F(\sigma^{-1}[k_s - \nu]) - F(\sigma^{-1}[k_{s-1} - \nu]),$$

dove  $F$  è la funzione di distribuzione cumulativa e  $\sigma$  è un parametro di scala che può essere modellato come  $\log(\sigma) = \boldsymbol{\alpha}' \mathbf{z}^{(1)}$ . Le risposte raggruppate o censurate continue si possono modellare in modo analogo vincolando i parametri soglia ad assumere valori uguali ai limiti degli intervalli di censura.

Nel caso di risposte conteggio (valori interi non negativi), il modello di risposta è specificato tipicamente tramite una funzione legame logaritmica ed una distribuzione di Poisson:

$$\ln(y | \mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) = \nu,$$

$$\text{e } f(y = s | \mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) = \frac{[\exp(\nu)]^s}{s!} \exp(-\exp(\nu)).$$

La distribuzione di Poisson si può utilizzare anche per modellare durate in tempo continuo<sup>101</sup>.

---

<sup>101</sup> Se si assume un modello esponenziale a passi (piecewise), con rischi costanti per gli intervalli di tempo, ciascuna durata osservata contribuisce alla verosimiglianza con un prodotto di termini aventi la forma della distribuzione di Poisson, cioè con un termine per ciascun intervallo che supera (Clayton, 1988). Se il rischio cambia ad ogni evento che si verifica nel tempo, equivale al modello di regressione di

Nel caso di risposte politomiche, graduatorie e confronti a coppie, tutte di tipo comparativo, si considera un modello a risposta latente avente la forma seguente:

$$y_s^* = v_s + \varepsilon_s,$$

dove  $s = 1, \dots, S$  indica modalità non ordinate. Per una risposta politomica:

$$y = y_s \quad \text{se} \quad y_s^* > y_\ell^*, \quad \forall \ell, \ell \neq s.$$

Se  $\varepsilon_s$  è specificato come una distribuzione di Gumbel (a valori estremi di tipo I),  $f(\varepsilon_s) = \exp(-\varepsilon_s - \exp(-\varepsilon_s))$ , la probabilità condizionata diventa un logit multinomiale:

$$f(y = y_s \mid \mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) = \exp(v_s) \left[ \sum_{\ell=1}^S \exp(v_\ell) \right]^{-1}.$$

Questa distribuzione può anche servire come elemento costitutivo per la distribuzione condizionata di graduatorie (Skrondal e Rabe-Hesketh, 2003a). Nei confronti a coppie, l'indicatore dicotomico di preferenza per coppie di alternative può essere modellato utilizzando una regressione probit o logit (Takane, 1987; Böckenholt, 2001).

Nel caso di risposte di tipo misto, si possono infine specificare funzioni legame e distribuzioni per le diverse tipologie di risposta. In generale, questo è il caso tipico in un modello ad equazioni strutturali con molte variabili latenti ed il modello di misura può richiedere funzioni legame e distribuzioni diverse.

### 2.9.3 Il modello a variabili latenti

Il modello strutturale per le variabili latenti assume la forma (Jöreskog, 1973; Muthén 1984):

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \mathbf{\Gamma}\mathbf{w} + \boldsymbol{\zeta},$$

---

Cox. Il problema della “overdispersion” può essere trattato con intercette casuali a livello 1

dove  $\mathbf{B}$  è una matrice  $M \times M$  di parametri strutturali per la relazione tra le variabili latenti  $\boldsymbol{\eta}$  ai diversi livelli,  $\mathbf{w}$  è un vettore di  $Q$  covariate,  $\boldsymbol{\Gamma}$  è una matrice  $M \times Q$  di parametri per la regressione delle variabili latenti sulle covariate e  $\boldsymbol{\zeta}$  è un vettore di  $M$  errori in cui ciascun elemento varia allo stesso livello del corrispondente elemento delle variabili latenti. La differenza tra questa formulazione, analoga a quella con un solo livello, è che le variabili latenti possono variare a livelli diversi. Ciascun elemento di  $\boldsymbol{\zeta}$  varia allo stesso livello del corrispondente elemento di  $\boldsymbol{\eta}$ .

#### 2.9.4 La log-verosimiglianza marginale

Quando le variabili latenti sono trattate come casuali ed i parametri come fissi<sup>102</sup>, l'inferenza è basata di solito sulla verosimiglianza marginale, la verosimiglianza dei dati condizionata alle variabili latenti, integrate (o sommate nel caso discreto) sulla distribuzione della variabile latente. Nel caso di variabili latenti continue, generalmente la verosimiglianza non ha una forma chiusa. In questo caso si ricorre ai metodi di integrazione approssimata, come i metodi numerici o l'integrazione Monte Carlo (verosimiglianza simulata).

Quindi, la verosimiglianza dei dati osservati è la verosimiglianza marginale a tutte le variabili latenti. Se  $\boldsymbol{\theta}$  è il vettore di tutti i parametri, inclusi i coefficienti di regressione  $\mathbf{B}$ , i coefficienti fattoriali  $\lambda_m^{(l)}$ ,  $m = 1, \dots, M_l$ ,  $l = 2, \dots, L$ , gli elementi non duplicati della matrice di covarianza  $\boldsymbol{\Sigma}_l$ , i parametri soglia,  $\kappa_s$ ,  $s = 2, \dots, S-1$  per le risposte ordinali, e i parametri  $\boldsymbol{\alpha}$  per modellare l'eteroschedasticità a livello 1. Inoltre, se  $\mathbf{y}_{(l)}$  è il vettore di variabili risposta e  $\mathbf{X}_{(l)}$  la matrice delle variabili esplicative per tutte le unità di livello 1 appartenenti ad una particolare unità a livello  $l$ , avente righe  $(\mathbf{x}', \mathbf{z}', \mathbf{z}^{(1)'}, \mathbf{w}')$ . Infine, siano  $\mathbf{y}$  ed  $\mathbf{X}$  rispettivamente i vettori delle variabili risposta e la matrice delle variabili esplicative per tutte le unità. La sostituzione del modello a variabili latenti nel modello di risposta restituisce la forma ridotta e definisce la distribuzione condizionata delle risposte date le variabili latenti e le variabili esplicative. La densità (o probabilità) condizionata corrispondente di una risposta riferita ad una

<sup>102</sup> Questo è l'approccio più comune, in contrasto all'approccio Bayesiano in cui le variabili latenti ed i parametri sono trattati come variabili casuali.

unità di livello 1 sarà indicata con  $f^{(l)}(\mathbf{y}_{(l)} | \mathbf{X}_{(l)}, \boldsymbol{\zeta}^{(2+)}; \boldsymbol{\theta})$ , dove  $\boldsymbol{\zeta}^{(l+)} = (\zeta^{(l)}, \dots, \zeta^{(L)})$ . La forma di questa funzione di densità è implicata dalle equazioni che definiscono il processo di risposta.

La densità normale multivariata delle variabili latenti a livello  $l$  è indicata invece con  $h^{(l)}(\boldsymbol{\zeta}^{(l)}; \boldsymbol{\theta})$ . La verosimiglianza marginale viene costruita in maniera ricorsiva. La densità condizionata delle risposte di una unità a livello  $l$ , condizionata alle variabili latenti a livello  $l+1$  o superiore, è data da:

$$f^{(l)}(\mathbf{y}_{(l)} | \mathbf{X}_{(l)}, \boldsymbol{\zeta}^{(l+1+)}; \boldsymbol{\theta}) = \int h^{(l)}(\boldsymbol{\zeta}^{(l)}; \boldsymbol{\theta}) \prod f^{(l-1)}(\mathbf{y}_{(l-1)} | \mathbf{X}_{(l-1)}, \boldsymbol{\zeta}^{(l+)}; \boldsymbol{\theta}) d\boldsymbol{\zeta}^{(l)},$$

dove il prodotto è esteso a tutte le  $l-1$  unità all'interno delle unità di livello  $l$ . La verosimiglianza marginale totale è il prodotto dei contributi di tutte le unità di livello più alto:

$$\ell(\boldsymbol{\theta}; \mathbf{y}; \mathbf{X}) = \prod f^{(L)}(\mathbf{y}_{(L)} | \mathbf{X}_{(L)}; \boldsymbol{\theta}).$$

La funzione di log-verosimiglianza marginale viene massimizzata con una integrazione numerica utilizzando un algoritmo Newton-Raphson. Per dati valori dei parametri, l'integrale multivariato sulle variabili latenti  $\boldsymbol{\zeta}^{(l)}$  viene valutato integrando su  $M_l$  variabili latenti indipendenti aventi una distribuzione normale standardizzata  $\mathbf{v}^{(l)}$  con  $\boldsymbol{\zeta}^{(l)} = \mathbf{C}_l \mathbf{v}^{(l)}$ , dove  $\mathbf{C}_l$  è la decomposizione di Cholesky<sup>103</sup> della matrice di covarianza  $\boldsymbol{\Sigma}_l$ . L'integrale può essere approssimato con la quadratura del prodotto Cartesiano<sup>104</sup> nel modo seguente:

<sup>103</sup> In algebra lineare la decomposizione di Cholesky è la fattorizzazione di una matrice hermitiana e definita positiva in una matrice triangolare inferiore e nella sua trasposta coniugata. Essa si può considerare come un caso speciale della più generale decomposizione LU. Il nome di questa decomposizione ricorda il matematico francese André-Louis Cholesky (1875-1918).

<sup>104</sup> Le variabili latenti sono perciò valutate ad una griglia rettangolare di punti. Un diverso numero di punti di quadratura  $R_m^{(l)}$  possono essere valutate per ciascuna variabile latente  $\nu_m^{(l)}$  richiedendo un totale di  $\prod_{m=1}^{M_l} R_m^{(l)}$  valutazioni dell'integrando. Una alternativa alla quadratura del prodotto cartesiano è la regola della quadratura sferica che è specificatamente designata per integrare densità multivariate normali (Stroud, 1971).

$$\begin{aligned}
& \int_{-\infty}^{+\infty} h^{(l)}(\boldsymbol{\zeta}^{(l)}; \boldsymbol{\theta}) \prod f^{(l-1)}(\mathbf{y}_{(l-1)} | \mathbf{X}_{(l-1)}, \boldsymbol{\zeta}^{(l+)}; \boldsymbol{\theta}) d\boldsymbol{\zeta}^{(l)} \\
&= \int_{-\infty}^{+\infty} \phi(\nu_{M_l}^{(l)}) \cdots \int_{-\infty}^{+\infty} \phi(\nu_1^{(l)}) \prod f^{(l-1)}(\mathbf{y}_{(l-1)} | \mathbf{X}_{(l-1)}, \mathbf{v}^{(l)}; \mathbf{v}^{([l+1]^+)}; \boldsymbol{\theta}) d\nu_1^{(l)} \cdots d\nu_{M_l}^{(l)} \\
&\approx \sum_{r_1=1}^{R_1^{(l)}} \pi_{r_1} \cdots \sum_{r_{M_l}=1}^{R_{M_l}^{(l)}} \pi_{r_{M_l}} \prod f^{(l-1)}(\mathbf{y}_{(l-1)} | \mathbf{X}_{(l-1)}, \alpha_{r_1}, \dots, \alpha_{r_{M_l}}, \mathbf{v}^{([l+1]^+)}; \boldsymbol{\theta}),
\end{aligned}$$

dove  $\mathbf{v}^{(l+)} = (\mathbf{v}^{(l)}, \dots, \mathbf{v}^{(L)})$ ,  $\phi(\cdot)$  è la funzione di densità normale standard mentre  $\pi_r$  e  $\alpha_r$  sono rispettivamente i pesi e la locazione della quadratura. Si possono utilizzare anche le regole di quadratura standard Gauss-Hermite<sup>105</sup> o, meglio di quadratura adattiva. Oltre a migliorare la stima dei parametri, la quadratura adattiva fornisce anche previsioni empiriche di Bayes delle variabili latenti migliori.

Un vantaggio della stima effettuata utilizzando la quadratura è che l'accuratezza può essere valutata confrontando soluzioni con diversi punti di quadratura. Dal momento che la quadratura adattiva è ricavata assumendo la normalità approssimata della densità a posteriori delle variabili latenti, tende a funzionare bene nel caso di risposte continue, conteggi e gruppi di grandi dimensioni, proprio laddove la quadratura standard spesso non va bene (Lesaffre e Spiessens, 2001).

---

<sup>105</sup> La quadratura approssima un integrale con una somma ponderata dell'integrando valutato ad una serie di valori o locazioni della variabile che deve essere integrata via. Le locazioni ed i pesi sono indicate come regole di quadratura.





## Capitolo 3

### Il modello per la valutazione dell'efficacia esterna del titolo di dottorato di ricerca

#### *Introduzione*

L'obiettivo di questo capitolo è quello di illustrare la costruzione di un modello ad equazioni strutturali multilivello con variabili latenti per la valutazione dell'efficacia esterna del titolo di dottorato di ricerca. Lo strumento di misura proposto che si intende analizzare è la variabile endogena definita "Efficacia esterna", indicata con (E). Tale variabile è stata ottenuta valutando il livello di soddisfazione per la condizione lavorativa attuale dei dottori di ricerca che hanno ottenuto il titolo nel 2008. Il giudizio è stato espresso ad un anno dal conseguimento del titolo ed è stato formulato su una scala a dieci modalità ordinate di risposta rispetto agli indicatori "coerenza con gli studi fatti", "utilizzo delle competenze acquisite" e "rispondenza ai propri interessi culturali" (capitolo 1, paragrafo 2). L'obiettivo del modello è quello di spiegare l'"Efficacia esterna" del titolo di dottorato in termini di una serie di variabili esplicative ed è stato stimato utilizzando i dati dell'indagine "Condizione attuale e prospettive occupazionali dei dottori di ricerca", commissionata dal CNVSU al Dipartimento di Statistica "G. Parenti" dell'Università degli Studi di Firenze.

L'organizzazione del capitolo è la seguente. La prima parte riguarda la definizione delle variabili esplicative latenti ed osservate (paragrafo 3.1), la scelta di ricorrere ad un modello ad equazioni strutturali multilivello (paragrafo 3.2) e la

descrizione del metodo di analisi a due livelli utilizzato (paragrafo 3.3). Successivamente, dopo aver descritto il sistema di relazioni ipotizzate (paragrafo 3.4) si giunge alla formalizzazione del modello (paragrafo 3.5), all'analisi dei risultati (paragrafo 3.6) ai limiti del modello proposto (paragrafo 3.7) e alle brevi conclusioni (paragrafo 3.8).

### **3.1 La selezione delle variabili esplicative**

L'efficacia esterna del titolo di dottorato di ricerca è probabilmente influenzata da una serie di variabili, alcune delle quali sono direttamente misurabili, con o senza errore, altre non sono misurabili.

Abbiamo già osservato che, ai fini della ricerca, gli aspetti più interessanti sono rappresentati dalle variabili non misurabili o latenti poiché si ritiene che gli effetti più importanti sull'”Efficacia esterna” possano derivare da elementi complessi in grado di rappresentare concetti multidimensionali. Tra questi sono considerati il livello di soddisfazione per l'attività lavorativa svolta, le aspettative per la propria carriera ed il grado di coinvolgimento nel proprio lavoro. In particolare, per quanto riguarda gli obiettivi della ricerca, sono considerate le variabili latenti “Soddisfazione”, “Aspettative” e “Partecipazione lavorativa”. Tali variabili sono definite in modo analogo a quanto fatto per l'”Efficacia esterna”, valutando il livello di soddisfazione riguardo ad una serie di aspetti che caratterizzano l'esperienza di lavoro attuale dei dottori di ricerca.

La soddisfazione lavorativa è un attributo molto complesso e coinvolge diversi aspetti della vita professionale di un individuo. In generale, la soddisfazione lavorativa è stata definita come uno stato emotivo piacevole derivante dalla valutazione del proprio lavoro (Locke, 1976), come una reazione affettiva al proprio lavoro (Cranny, Smith e Stone, 1992) e come un atteggiamento verso il proprio lavoro (Brief, 1998). Weiss (2002) sostiene che la soddisfazione sul lavoro è un'opinione e sottolinea che la valutazione della propria attività lavorativa si forma anche tenendo conto delle emozioni e dei sentimenti che, invece, dovrebbero rimanere chiaramente separati.

Il metodo più comune per misurare la soddisfazione lavorativa è quello di valutare una serie di aspetti che la caratterizzano (varietà dei compiti svolti,

responsabilità, livello di retribuzione, opportunità di promozione, rapporti con i colleghi ed altri ancora) utilizzando comuni scale di misurazione degli atteggiamenti. Questo è anche il metodo utilizzato per la ricerca. In particolare, per questo scopo sono state utilizzate alcune risposte alla domanda “A30” (cfr. prospetto successivo), alle quali i dottori potevano esprimere il proprio giudizio su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo):

*A 30 - Può esprimere su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo) il suo livello di soddisfazione per i seguenti aspetti del lavoro attuale?*

	Aspetti dell'attività lavorativa	Per niente		Moltissimo
1	Livello di retribuzione attuale	1	...	10
2	Prospettive di guadagno	1	...	10
3	Possibilità di carriera	1	...	10
4	Stabilità/sicurezza sul lavoro	1	...	10
5	Coerenza con studi fatti	1	...	10
6	Utilizzo delle competenze acquisite	1	...	10
7	Acquisizione di professionalità	1	...	10
8	Rispondenza ai propri interessi culturali	1	...	10
9	Indipendenza e autonomia sul lavoro	1	...	10
10	Coinvolgimento nei processi decisionali	1	...	10
11	Flessibilità dell'orario di lavoro	1	...	10
12	Tempo libero	1	...	10
13	Localizzazione sede di lavoro	1	...	10
14	Rapporti coi colleghi	1	...	10
15	Rispetto alle aspettative iniziali	1	...	10
16	Rispetto alle mansioni svolte	1	...	10
17	Rispetto al titolo di Dottore di Ricerca	1	...	10
18	E nel suo complesso	1	...	10

Per quanto riguarda le variabili esplicative osservate, il criterio che ha guidato la scelta è stato quello di considerare gli aspetti che hanno caratterizzato l'esperienza del dottorato, ad esempio se il dottorando disponeva di una borsa di studio, se ha trascorso un periodo all'estero, se svolgeva un'attività lavorativa ecc.. Non sono invece state prese in considerazione le variabili riferite all'attività post-dottorato e al tipo di lavoro svolto perché sono esse stesse influenzate dall'esperienza del dottorato.

### *3.1.1 La misura della soddisfazione*

In questa sede, alla soddisfazione lavorativa viene attribuita la valenza di variabile non osservata o latente la cui esistenza determina associazione tra i valori osservati degli indicatori considerati. Tale variabile è stata costruita considerando alcune risposte alla domanda “A30” ed è stata definita “Soddisfazione” ed indicata con (S). In particolare

sono stati considerati gli aspetti ritenuti, quantomeno in via di prima approssimazione, più rilevanti, cioè, la soddisfazione:

- rispetto alle aspettative iniziali;
- rispetto al titolo di dottore di ricerca;
- nel suo complesso.

Tali aspetti si ritengono correlati positivamente alla “Soddisfazione”. La scelta di utilizzare questi tre indicatori si spiega con l’intenzione di valutare non solo la soddisfazione per l’attività svolta ma anche alcuni aspetti che la caratterizzano, considerando rilevanti sia la formazione ricevuta nel corso di dottorato che le aspettative iniziali. Quindi, anche in questo caso, analogamente a quanto fatto per l’”Efficacia esterna”, sono preferiti gli aspetti di tipo qualitativo. Per questo motivo, non è stato considerato il grado di soddisfazione rispetto al livello di retribuzione attuale perché si ritiene indipendente dal grado di soddisfazione rispetto alle caratteristiche dell’attività svolta. Tuttavia, poiché si pensa che il grado di soddisfazione rispetto al livello di retribuzione sia un aspetto non trascurabile della vita professionale dei dottori, questo sarà considerato a parte come variabile esplicativa.

La relazione ipotizzata tra la variabile latente “Soddisfazione” e gli indicatori osservati (modello di misura) si può rappresentare graficamente nel modo seguente (Figura 3.1).

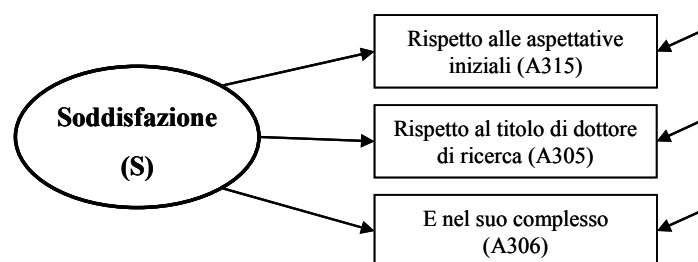


Figura 3.1. Relazione tra la variabile latente “Soddisfazione” e gli indicatori osservati

### 3.1.2 La misura delle “Aspettative” e della “Partecipazione lavorativa”

La misurazione delle “Aspettative” (A) e della “Partecipazione lavorativa” (PL) segue la stessa logica utilizzata per l’”Efficacia esterna” e per la “Soddisfazione”, definendo

una variabile latente la cui esistenza determina associazione tra i valori osservati degli indicatori considerati. Anche in questo caso, gli indicatori osservati sono stati scelti tra le risposte alla domanda “A30”.

Le “Aspettative” sono state misurate considerando la soddisfazione espressa rispetto:

- alle prospettive di guadagno;
- alle possibilità di carriera;
- alla stabilità/sicurezza del lavoro.

Tali aspetti si ritengono correlati positivamente alle “Aspettative”. La scelta di tali indicatori si spiega con l’importanza che questi assumono rispetto alle attese che i dottori di ricerca hanno nei confronti della propria vita professionale. In particolare, oltre alle prospettive di guadagno e alle possibilità di carriera, che sono rilevanti in qualunque momento della propria carriera, si ritiene di valutare anche la stabilità/sicurezza del lavoro. Questo aspetto appare molto importante soprattutto in un periodo, come quello attuale, caratterizzato dalla prevalenza di forme contrattuali di tipo precario che di solito sono riservate a coloro che entrano per la prima volta nel mondo del lavoro.

La relazione tra gli indicatori osservati e la variabile latente “Aspettative” può essere rappresentata nel modo seguente (Figura 3.2):

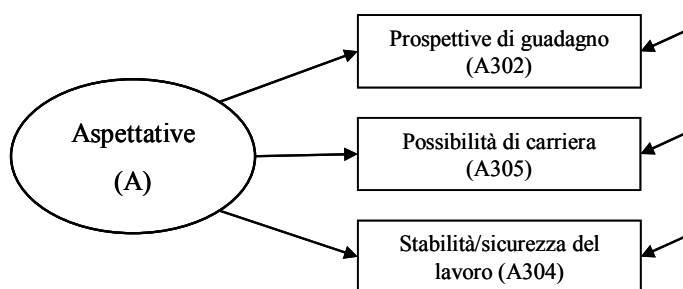


Figura 3.2. Relazione tra la variabile latente “Aspettative” e gli indicatori osservati

La “Partecipazione lavorativa” è stata misurata considerando i livelli di soddisfazione espressi relativamente:

- all'acquisizione di professionalità;
- all'indipendenza e autonomia sul lavoro;
- al coinvolgimento nei processi decisionali.

L'indipendenza e autonomia sul lavoro ed il coinvolgimento nei processi decisionali sono aspetti che indicano piuttosto chiaramente una condivisione partecipata delle attività che qualificano la propria attività lavorativa. Per quanto riguarda l'acquisizione di professionalità, si ritiene che possa rappresentare un aspetto della partecipazione lavorativa in quanto svolgere un lavoro altamente professionalizzante comporta probabilmente un maggior coinvolgimento, più interesse per le mansioni svolte e quindi anche un più alto livello di condivisione delle attività. Gli indicatori si considerano correlati positivamente alla "Partecipazione lavorativa".

La relazione tra gli indicatori osservati e la variabile latente "Partecipazione lavorativa" può essere rappresentata nel modo seguente (Figura 3.3):

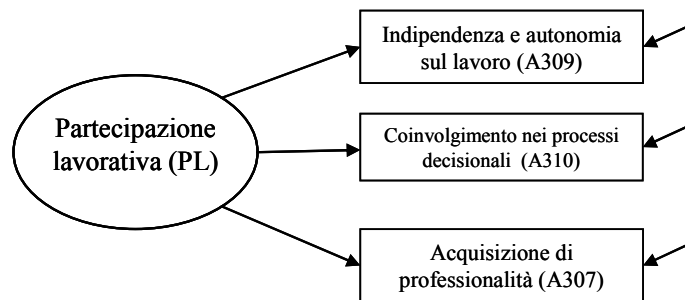


Figura 3.3. Relazione tra la variabile latente "Partecipazione lavorativa" e gli indicatori osservati

### 3.2 La scelta del modello

L'analisi descrittiva preliminare delle risposte (Appendice C) non è soddisfacente. Il fenomeno che si vuole analizzare è piuttosto complesso perché coinvolge diverse variabili latenti e variabili osservate. Un modello statistico consente di descrivere la struttura delle relazioni tra tutte le variabili considerate e di formalizzare una teoria in termini semplificati. I dati sono stati analizzati con un modello ad equazioni strutturali a due livelli a variabili latenti, considerando i dottori di ricerca come unità di primo livello ed i corsi di dottorato come unità di secondo livello. Il modello è stato scelto per i vantaggi dei modelli multilivello e per i vantaggi dei modelli ad equazioni strutturali.

Per quanto riguarda i vantaggi dei modelli multilivello, sui quali si è già avuto modo di soffermare l'attenzione (Capitolo 2), la scelta di un'analisi disaggregata rispetto all'analisi aggregata si giustifica innanzitutto perché consente di esplicitare la struttura gerarchica presente nei dati. Nel caso specifico, i dottori di ricerca sono raggruppati in corsi di dottorato che sono compresi in aree disciplinari, in atenei e magari in altre unità di livello ancora superiore. In conseguenza di tale struttura, è possibile che le unità statistiche appartenenti allo stesso gruppo per ciascun livello della gerarchia, se condividono fattori non osservati, abbiano valori della variabile risposta correlati. In questo caso, l'analisi multilivello, oltre a spiegare la struttura di variabilità complessa, analizzando la variabilità associata a ciascun livello, consente di ottenere stime corrette dei parametri e degli errori standard.

La scelta di un'analisi multilivello si giustifica anche per la possibilità di analizzare in un unico modello statistico variabili misurate a livelli diversi, riferite rispettivamente alle unità di primo e di secondo livello. Nel caso specifico questo significa che si possono valutare gli effetti esercitati sulla variabile endogena dalle variabili individuali (età, sesso, tipologia di attività lavorativa svolta, ecc.) e dalle variabili di contesto (tipologia dei corsi, aree disciplinari e variabili ottenute per aggregazione di variabili osservate a livello individuale).

L'analisi è limitata a due soli livelli di aggregazione principalmente perché si ritiene che gli effetti maggiori sull'efficacia esterna siano esercitati dalle unità poste al livello gerarchico immediatamente superiore (corsi di dottorato) a quello delle unità elementari considerate (dottori di ricerca). Inoltre, l'analisi a due livelli è preferibile anche perché consente un'interpretazione più semplice dei risultati.

Per quanto riguarda i vantaggi dei modelli ad equazioni strutturali, la scelta si giustifica innanzitutto per la possibilità di analizzare un sistema multivariato di relazioni tra variabili (esogene ed endogene) attraverso un sistema di equazioni. Questo permette di modellare la variabile dipendente in un'equazione come variabile indipendente in un'altra equazione e di scomporre gli effetti esercitati da una variabile sulle altre variabili, in effetti, diretti ed indiretti. Nel caso specifico, questo si traduce nella possibilità di modellare, ad esempio, la variabile "Soddisfazione" come dipendente quando le variabili esplicative sono le "Aspettative" e la "Partecipazione lavorativa" e come variabile esplicativa quando la variabile dipendente è l'"Efficacia esterna".

Inoltre, questo consente di valutare l'effetto totale che la variabile "Aspettative" esercita sull'"Efficacia esterna", distinguendo l'effetto diretto dall'effetto indiretto esercitato attraverso la "Soddisfazione".

La scelta di un modello ad equazioni strutturali si giustifica anche per la possibilità di modellare variabili latenti. In questo modo è possibile analizzare relazioni tra variabili che non possono essere misurate perfettamente, o perché non osservabili, e quindi misurate da una serie di indicatori osservati non completamente adeguati rispetto a ciò che si vuole misurare, o perché affette da errore di misura. Nel caso specifico, le variabili latenti che si vogliono analizzare, "Efficacia esterna", "Soddisfazione", "Aspettative" e "Partecipazione lavorativa" sono definite attraverso una serie di indicatori misurabili che si ipotizzano a queste correlati. Poiché è ragionevole ritenere che tali indicatori, per quanto appropriati, non siano in grado di misurare perfettamente le variabili latenti, modellare anche l'errore di misura consente di ottenere stime più accurate dei coefficienti che definiscono le relazioni tra le variabili analizzate.

### 3.3 Le caratteristiche del modello

Il modello utilizzato è un modello ad equazioni strutturali è stato ottenuto seguendo l'approccio tradizionale che prevede la formulazione di modelli separati per le matrici di covarianza entro i gruppi e tra i gruppi (capitolo 2, paragrafo 2.8.1). Questo approccio si utilizza di solito quando le variabili risposta osservate sono continue o categoriche modellate da variabili latenti continue sottostanti.

Il modello prevede, come già sottolineato, una struttura dati a due livelli in cui sono osservate  $N$  unità statistiche (dottori di ricerca) suddivise in  $J$  gruppi (corsi di dottorato) per la quale si assume variabilità nelle risposte a livello individuale e a livello di gruppo. In questo caso, il vettore di variabili osservate riferite all'unità  $i$  appartenente al gruppo  $j$ , indicato con  $\mathbf{y}_{ij}$ , si può scomporre in una componente che varia tra i gruppi,  $\mathbf{y}_b = \bar{\mathbf{y}}_j$ , e una componente che varia entro i gruppi,  $\mathbf{y}_w = \mathbf{y}_{ij} - \bar{\mathbf{y}}_j$ . Le due componenti sono ortogonali e additive:

$$\mathbf{y}_{ij} = \mathbf{y}_b + \mathbf{y}_w.$$



La scomposizione del vettore delle variabili osservate può essere utilizzata per calcolare una matrice di covarianza tra i gruppi, indicata con  $\Sigma_B$ , e una matrice di covarianza entro i gruppi, indicata con  $\Sigma_W$ . Tali matrici rappresentano, rispettivamente, la matrice di covarianza riferita agli scarti tra le medie di gruppo e la media generale e la matrice di covarianza riferita agli scarti delle unità dalle medie di ciascun gruppo e sono a loro volta ortogonali e additive:

$$\Sigma_T = \Sigma_B + \Sigma_W.$$

Per la stima di tali modelli sono stati proposti diversi metodi, la maggior parte dei quali prevedono stimatori di massima verosimiglianza e sono adatti quando le variabili risposta osservate sono continue (capitolo 2, paragrafi 2.8.3 e 2.8.4).

Nel caso specifico, le variabili risposta osservate sono di natura categorica con modalità ordinate. La presenza di variabili risposta osservate categoriche aggiunge notevole complessità alla stima dei modelli ad equazioni strutturali multilivello poiché non esiste una soluzione in forma chiusa per la funzione di verosimiglianza marginale (Rabe-Hesketh, Skrondal, e Pickles, 2005) che deve essere approssimata da una procedura di calcolo intensiva ad integrazione numerica.

Per la stima del modello è stato quindi utilizzato l'approccio seguito da Asparouhov e Muthén (2007) (capitolo 2, paragrafo 2.8.6) che, a differenza dei metodi precedenti, si basa sul metodo di stima ad informazione limitata e sulla funzione di adattamento dei minimi quadrati ponderati (capitolo 2, paragrafo 2.4.6). Tale approccio consente di stimare modelli ad equazioni strutturali a due livelli con variabili risposta osservate binarie, categoriche ordinali, censurate, continue o combinazioni tra queste. Tale stimatore è implementato nel software *Mplus* (versione 5 e successive) dove è indicato con l'acronimo "WLSM" o "WLSMV"<sup>106</sup> e restituisce stime efficienti dei parametri e test di adattamento del modello corretti in presenza di variabili risposta osservate categoriche.

---

<sup>106</sup> La differenza tra i due è solo nel calcolo degli indici robusti di adattamento del modello, definiti rispettivamente "*mean-adjusted*" e "*mean and variance-adjusted*" mentre le stime dei parametri sono le stesse.

Tale approccio è stato preferito perché si può utilizzare per stimare modelli ad equazioni strutturali con un numero qualsiasi di effetti casuali senza che questo comporti un allungamento dei tempi di calcolo o comprometta la precisione delle stime. I limiti principali sono rappresentati dalla possibilità di stimare solo modelli ad intercetta casuale e dalla impossibilità di prevedere interazioni tra variabili a livelli diversi.

In particolare, questo approccio è stato preferito a quello di Rabe-Hesketh, Skrondal e Pickles (capitolo 2, paragrafo 2.9) che consente di analizzare modelli ad equazioni strutturali generalizzati multilivello. I modelli ad equazioni strutturali generalizzati multilivello non prevedono la stima di matrici di covarianza separate tra i gruppi ed entro i gruppi e consentono di modellare una più ampia gamma di variabili risposta. Tuttavia, la maggiore flessibilità che caratterizza questo approccio, data anche dalla possibilità trattare più di due livelli di analisi, di prevedere intercette e coefficienti casuali e interazioni tra variabili a livelli diversi, non pare compensata dalla maggiore complessità della stima che, basandosi sulla funzione di verosimiglianza marginale, risulta molto pesante e quasi inapplicabile quando le variabili risposta osservate sono categoriche ed il modello prevede molteplici variabili latenti.

### **3.4 Il sistema di relazioni ipotizzate**

Come già sottolineato, il modello che si intende stimare è un modello ad equazioni strutturali a due livelli in cui le unità di primo livello sono i dottori di ricerca e le unità di secondo livello sono i corsi di dottorato. L'“Efficacia esterna” è la variabile latente endogena che si vuole spiegare ed è misurata a livello individuale (dottori di ricerca) e a livello di gruppo (corsi di dottorato). In particolare, lo scopo è quello di capire quali sono i fattori che la influenzano a livello individuale e a livello di gruppo, tenendo sotto controllo le caratteristiche delle unità di primo e di secondo livello, riassunte nelle variabili esplicative osservate considerate.

#### *3.4.1 Il sistema di relazioni di primo livello*

Il primo livello di osservazione prevede l'analisi delle relazioni esercitate sulla variabile latente “Efficacia esterna” (E) dalle variabili latenti “Soddisfazione” (S),

“Partecipazione lavorativa” (PL) ed “Aspettative” (A) tenendo sotto controllo le variabili osservate “Durante il Corso di Dottorato in cui ha conseguito il titolo di Dottore di Ricerca Lei è stato titolare di borsa di studio?” (Z02), “Durante il Corso di Dottorato ha compiuto periodi di studio all'estero?” (Z05), “Durante il Corso di Dottorato ha svolto una qualche attività lavorativa (inclusi co.co.co e co-co.co.pro.) (Z06), “Grado di soddisfazione per il livello di retribuzione attuale” (A301), “genere” ed “età”.

In particolare, si ipotizza che, per ciascun dottore (unità di primo livello), essere soddisfatti per la condizione lavorativa attuale “rispetto alle mansioni svolte”, “rispetto al titolo di dottorato di ricerca” e “nel suo complesso” abbia un effetto positivo sull’“Efficacia esterna”, cioè sulla possibilità che l’attività sia coerente con gli studi svolti”, che le “competenze acquisite siano effettivamente utilizzate” e che l’attività lavorativa sia in linea con gli interessi culturali” dei dottori. Più precisamente, l’“Efficacia esterna” si ritiene maggiore nei casi in cui i dottori sono più soddisfatti per l’attività svolta. Questa ipotesi si traduce in una relazione diretta tra la variabile latente “Soddisfazione” e la variabile latente “Efficacia esterna”.

Inoltre, si ipotizza che abbiano effetti positivi sull’“Efficacia esterna” anche il “coinvolgimento nei processi decisionali”, la possibilità di “acquisire professionalità” ed avere “indipendenza ed autonomia sul lavoro”. Questa seconda ipotesi si traduce in una relazione diretta tra la variabile latente “Partecipazione lavorativa” e la variabile latente “Efficacia esterna”. Tuttavia, è ragionevole ritenere che la “Partecipazione lavorativa” non si limiti ad avere un effetto diretto sull’“Efficacia esterna” ma eserciti anche un effetto indiretto, attraverso la “Soddisfazione”. In particolare, si ritiene che una parte dell’effetto che la “Soddisfazione” esercita sull’“Efficacia esterna” sia dovuto proprio alla “Partecipazione lavorativa” nei confronti della quale la “Soddisfazione” agisce da mediatore.

Infine, si ritiene che anche le “Aspettative” possano esercitare un effetto sulla variabile latente “Efficacia esterna”. In particolare, si ipotizza che le “Aspettative” abbiano un effetto diretto ed un effetto indiretto sull’“Efficacia esterna”, dove l’effetto indiretto è esercitato, anche in questo caso, attraverso la “Soddisfazione”. Infatti, si ritiene che le “possibilità di carriera”, le “prospettive di guadagno” e la “stabilità nei confronti dell’attività lavorativa” possano produrre effetti diretti sulla “Soddisfazione”

attuale che, a sua volta, si riflette sull'”Efficacia esterna”. Più incerto è invece il segno dell'effetto diretto che si ipotizza esistere tra “Aspettative” ed ”Efficacia esterna”. Infatti, ad avere aspettative positive per la propria carriera possono essere sia coloro per i quali il titolo è risultato “efficace” nei confronti dell'attività lavorativa attuale sia coloro per i quali il titolo di dottorato non è risultato “efficace” ma ritengono che lo sarà, considerato il breve periodo di tempo trascorso dal conseguimento del titolo.

Il modello si completa considerando le caratteristiche delle unità di primo livello che sono necessarie per depurare le relazioni ipotizzate tra le variabili latenti dalla presenza di eventuali altri effetti. In particolare, si ritiene innanzitutto che sia necessario considerare l'effetto del genere e dell'età sull'”Efficacia esterna”. L'effetto di genere si giustifica soprattutto per le diverse opportunità lavorative che di solito i maschi hanno rispetto alle femmine. L'effetto dell'età è dovuto all'intento di distinguere i dottori che si sono iscritti ad un corso di dottorato subito dopo aver conseguito la laurea (o comunque entro i trenta anni di età) da coloro che si sono iscritti più tardi. Per questo, la variabile età è stata trattata come binaria, scegliendo come soglia trentadue anni, il che significa distinguere coloro che hanno conseguito il titolo entro i trentuno anni di età dagli altri.

Inoltre, si suppone che il “Grado di soddisfazione per il livello di retribuzione attuale” possa agire sulla “Soddisfazione” e sulle “Aspettative”. In particolare, si ritiene che essere soddisfatti per la propria retribuzione attuale possa esercitare effetti positivi in termini di “possibilità di carriera”, di “prospettive di guadagno” e magari anche di “stabilità e sicurezza del lavoro”. Questo perché si ritiene che un livello di retribuzione soddisfacente già entro un anno dal conseguimento del titolo di dottorato sia associato ad una posizione lavorativa di maggiore prestigio, indipendentemente dal tipo di attività svolta, e quindi possa provocare aspettative positive di carriera e di retribuzione per il prossimo futuro. Inoltre, è ragionevole ipotizzare che essere soddisfatti per la propria retribuzione attuale comporti anche un effetto diretto sulla “Soddisfazione” e quindi uno indiretto sull'”Efficacia esterna”. Invece, si ritiene che il “Grado di soddisfazione per il livello di retribuzione attuale” non agisca direttamente sull'”Efficacia esterna”. Infatti, si pensa che avere un livello retributivo soddisfacente non comporti necessariamente un'”attività coerente con gli studi svolti”, in cui le “competenze acquisite siano effettivamente utilizzate” ed “in linea con gli interessi culturali” dei dottori, soprattutto

se si considera che le caratteristiche precedenti, per un dottore di ricerca, sono più facilmente associate ad un'attività di ricerca che normalmente non prevede remunerazioni particolarmente elevate.

Le altre variabili esplicative considerate sono le variabili osservate “Durante il Corso di Dottorato in cui ha conseguito il titolo di Dottore di Ricerca Lei è stato titolare di borsa di studio?” (Z02), “Durante il Corso di Dottorato ha compiuto periodi di studio all'estero?” (Z05) e “Durante il Corso di Dottorato ha svolto una qualche attività lavorativa (inclusi co.co.co e co-co.co.pro.) (Z06). Tali variabili sono state poste in relazione con l'”Efficacia esterna” per tenere sotto controllo gli effetti che queste possono esercitare nel determinare le risposte degli indicatori di efficacia considerati.

Il sistema complesso di relazioni descritte, porta ragionevolmente a pensare ad un sistema di equazioni ricorsivo<sup>107</sup>, in cui la prima variabile latente compare nelle equazioni successive come variabile esplicativa.

In conclusione, il modello riferito alle unità di primo livello si concentra sugli effetti che le variabili latenti “Soddisfazione”, “Partecipazione lavorativa” e “Aspettative” hanno nei confronti della variabile latente endogena “Efficacia esterna”. In particolare, il modello evidenzia e vuole verificare il ruolo fondamentale nei confronti dell'”Efficacia esterna” svolto dalla “Soddisfazione” che, oltre ad esercitare un effetto diretto, costituisce un mediatore per gli effetti indiretti esercitati rispettivamente dalla “Partecipazione lavorativa” e dalle “Aspettative”.

### *3.4.2 Il sistema di relazioni di secondo livello*

Il secondo livello di osservazione prevede l'analisi della variabile latente endogena “Efficacia esterna del corso” (EB). La variabile è misurata a livello di corso di dottorato utilizzando gli stessi indicatori di efficacia considerati per le unità di primo livello, vale a dire “Coerenza con gli studi fatti”(A305), “Utilizzo delle competenze acquisite” (A306) e “Rispondenza ai propri interessi culturali” (A308). In pratica, a parità di altre condizioni, un corso di dottorato di ricerca si ritiene efficace rispetto alla condizione

---

<sup>107</sup> In genere, ricorsivo è un sistema di equazioni che possono essere ordinate in modo tale da poter essere risolto in maniera sequenziale. In questo senso, è possibile usare l'informazione della prima equazione per risolvere le equazioni successive, così come le altre equazioni servono per risolvere le equazioni che le seguono, dato che ogni variabile endogena dipende, oltre che dalle variabili esogene, dalle altre variabili endogene che la precedono. Nei modelli ad equazioni strutturali il termine ricorsivo indica un modello in cui si ipotizza un rapporto di dipendenza ad un'unica direzione.

lavorativa dei dottori se l'attività svolta dai dottori di ricerca che hanno frequentato il corso è coerente con gli studi svolti, se le competenze acquisite sono effettivamente utilizzate e se l'attività lavorativa è in linea con gli interessi culturali dei dottori. In particolare, l'”Efficacia esterna del corso” è ritenuta maggiore in quei corsi in cui tali indicatori assumono valori più elevati.

Le variabili esplicative dell'” Efficacia esterna del corso” non sono numerose, anzi, sono decisamente poche, principalmente per l'impossibilità di conoscere le caratteristiche specifiche dei corsi di dottorato, che probabilmente rendono i corsi stessi molto eterogenei tra loro.

Tra le variabili di cui si dispone, è stato considerato innanzitutto il parere dei dottori circa la possibilità di una nuova iscrizione allo stesso corso di dottorato. In particolare, questo è stato ricavato dalla domanda “Se potesse tornare indietro, si iscriverebbe di nuovo ad un corso di dottorato?” (Z04), ed i suoi valori sono stati ottenuti aggregando per corso di dottorato la percentuale di risposte “Sì, allo stesso corso nella stessa sede universitaria”. In pratica, si pensa che tale giudizio possa rappresentare una misura di come i dottori percepiscono la qualità del corso di dottorato frequentato. Il limite maggiore di questo indicatore è che, basandosi su un giudizio espresso ad un anno circa dal conseguimento del titolo di dottore di ricerca, il suo valore potrebbe non dipendere solo dal livello di qualità del corso ma anche dalle aspettative individuali e dai risultati conseguiti sul mercato del lavoro, ormai definiti al momento dell'intervista. Per questo motivo, l'assunzione di esogeneità delle variabili esogene potrebbe non essere rispettata (capitolo 2, paragrafo 2.4.1).

Inoltre, si è provato a misurare la “Qualità del corso” anche con una variabile latente composita, ottenuta considerando gli indicatori presenti alla domanda Z03, in cui si chiedeva ai dottori di ricerca di esprimere il grado di soddisfazione rispetto ad una serie di aspetti della formazione ricevuta. Gli aspetti considerati per la variabile latente “Qualità del corso” sono stati il “contenuto degli insegnamenti”(Z301), la “Qualità della didattica del personale docente” (Z302), la “Qualità scientifica di professori/ricercatori” (Z303), lo “Sviluppo/addestramento della capacità di fare ricerca” (Z304) e la “Possibilità di pubblicare (articoli, volumi, ...)”. Per ciascun aspetto, il giudizio è stato misurato con una variabile categorica suddivisa in dieci modalità ordinate (1 = “per niente” e 10 = “moltissimo”). Analogamente a quanto ipotizzato per gli altri indicatori

misurati sulla medesima scala, ognuna di queste variabili risposta è stata considerata come un indicatore osservato di una variabile latente sottostante. L'ipotesi è che la "qualità" dei corsi sia migliore nei corsi in cui i diversi indicatori osservati assumono valori più elevati. Purtroppo, anche per la variabile latente "Qualità del corso" potrebbe non essere rispettata l'ipotesi di esogeneità.

Per provare a comprendere se l'"Efficacia esterna del corso" dipende dalle caratteristiche dei corsi, sono state considerate anche le variabili esplicative dicotomiche indicative delle diverse aree disciplinari "Area.." (cfr. Schema sotto riportato) e una variabile che distingue i corsi delle università private da quelli delle università pubbliche "UniPriv" (0 = università pubblica; 1 = università privata). In particolare, si ritiene che le aree disciplinari possano spiegare eventuali differenze nell'"Efficacia esterna del corso". L'Area 1 (Scienze matematiche, informatiche e fisiche) è stata presa a riferimento perciò i valori dei coefficienti riferiti alle altre aree indicano eventuali differenze rispetto a questa.

*Schema delle macro aree disciplinari*

Area 1	Scienze matematiche, informatiche e fisiche
Area 2	Scienze chimiche e della terra
Area 3	Scienze biologiche, mediche, agrarie e veterinarie
Area 4	Ingegneria civile ed Architettura, Ingegneria industriale e dell'informazione
Area 5	Scienze dell'antichità, filologico-letterarie e storico-artistiche, Scienze storiche, filosofiche, pedagogiche e psicologiche
Area 6	Scienze giuridiche
Area 7	Scienze economiche e statistiche, Scienze politiche e sociali

La natura dell'università (pubblica o privata) è stata considerata perché si ritiene che possa qualificare un corso di dottorato, anche se a rigore non rappresenta una caratteristica dell'unità di secondo livello (corso di dottorato) ma eventualmente dell'unità di terzo livello (ateneo) in cui questa è compresa.

### 3.4.3 La rappresentazione grafica del modello proposto

La rappresentazione grafica delle relazioni segue la proposta di Muthén e Muthén (2004) e distingue il modello entro i gruppi (within), rappresentato in basso, dal modello tra i gruppi (between), rappresentato in alto (Figura 3.4).

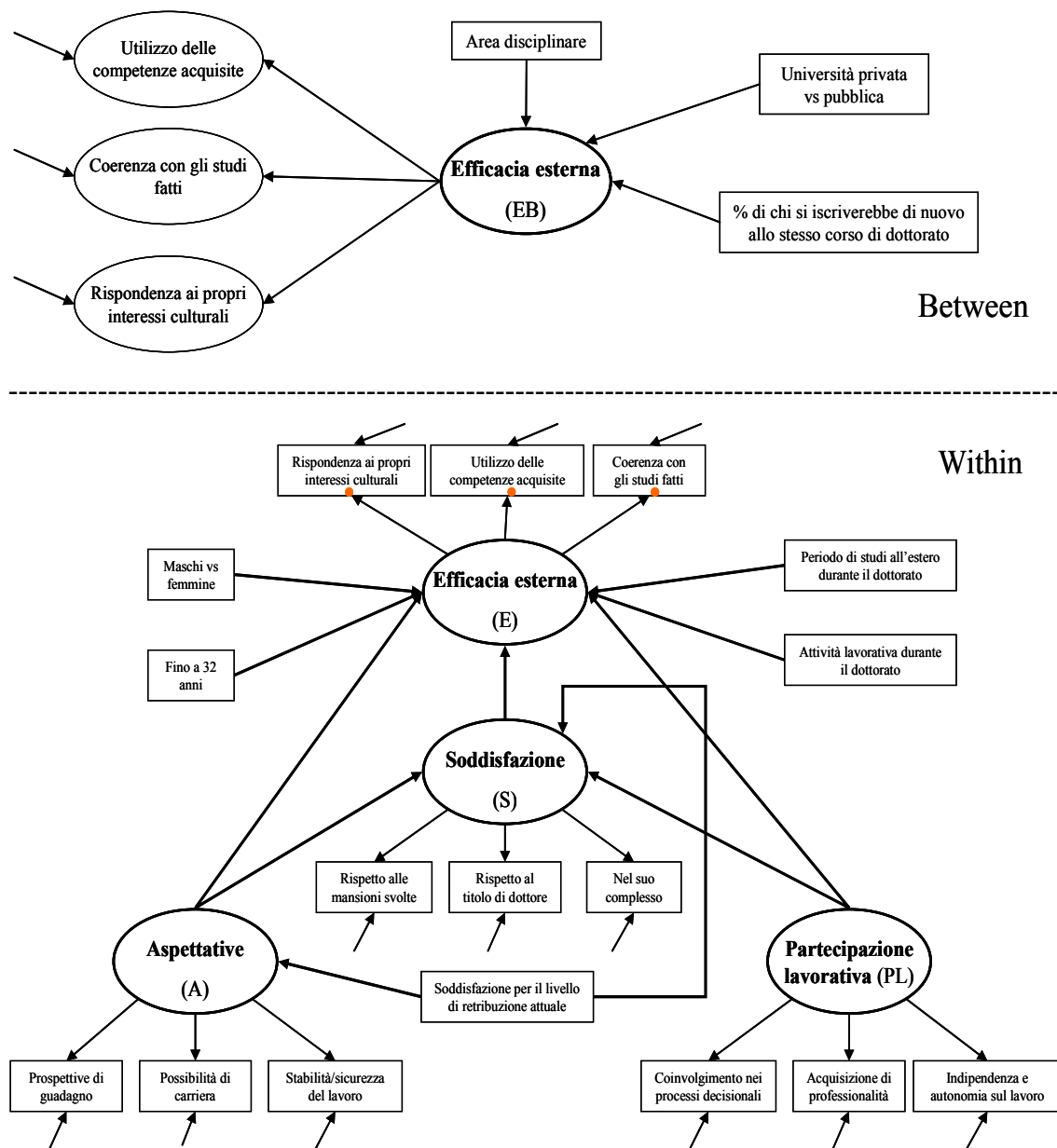


Figura 3.4. Rappresentazione grafica del modello proposto.

La parte entro i gruppi mostra il modello di misura, le relazioni ipotizzate tra le variabili latenti e le relazioni tra le variabili latenti e le variabili esplicative osservate, tutte modellate a livello delle unità individuali (dottori di ricerca). Le variabili osservate, di



primo e di secondo livello<sup>108</sup>, sono comprese in rettangoli mentre le variabili latenti sono comprese in ellissi. Nel modello di misura, i punti solidi in corrispondenza dell'incontro tra le frecce direzionali che partono dalla variabile latente e le variabili osservate stanno ad indicare che la relazione tra le variabili latenti e gli indicatori osservati sono casuali a livello di gruppo. Nella parte tra i gruppi del modello, le intercette casuali sono rappresentate come ellissi poiché sono variabili latenti, modellate a livello di gruppo.

La versione finale del modello posta a verifica è quella riportata nella figura 3.4 e si differenzia leggermente rispetto a quanto previsto nelle ipotesi iniziali (paragrafo 3.4.2). La parte entro i gruppi del modello non comprende la variabile osservata “Durante il Corso di Dottorato in cui ha conseguito il titolo di Dottore di Ricerca Lei è stato titolare di borsa di studio?” (Z02) poiché non è risultata significativa. La parte tra i gruppi non comprende invece la variabile latente “Qualità del corso” (capitolo 3, paragrafo 3.4.2). Tale variabile è stata esclusa dopo aver verificato che il suo contributo alla spiegazione della variabile “Efficacia esterna del corso” non è significativo se considerato insieme alla variabile “Se potesse tornare indietro, si iscriverebbe di nuovo ad un corso di dottorato?” (Z04). Il modello così ottenuto risulta notevolmente semplificato ed i tempi di stima sensibilmente ridotti, considerato che la variabile latente “Qualità del corso” è piuttosto complessa da stimare.

### **3.5 La formalizzazione del modello**

Il modello prevede una formulazione separata, ottenuta distinguendo la misurazione delle variabili latenti dalle relazioni ipotizzate tra le variabili esogene ed endogene. In particolare, il modello ad equazioni strutturali a due livelli è stato specificato da un sistema di equazioni per ciascuna variabile latente (modello di misura) e da un sistema di equazioni per il modello che definisce le relazioni tra le variabili latenti e tra le variabili latenti e le variabili esplicative osservate.

---

<sup>108</sup> Questa convenzione è rispettata anche per le variabili di secondo livello ottenute per aggregazione di variabili di primo livello.

### 3.5.1 La misurazione delle variabili latenti

Il modello di misura stabilisce la relazione tra le variabili latenti e le variabili risposta osservate (indicatori misurabili) ed è specificato da un modello fattoriale confermativo per ciascuna variabile latente considerata (capitolo 2, paragrafo 2.1.2).

Gli indicatori osservati sono misurati con variabili categoriche ordinali quindi il modello di misura è definito per le variabili latenti continue sottostanti, che si ipotizzano distribuite normalmente (capitolo 2, paragrafo 2.3.1). Per rappresentare la struttura gerarchica a due livelli, le variabili risposta osservate sono indicate con  $y_{ij}$ , mentre le corrispondenti variabili latenti sottostanti sono indicate con  $y_{ij}^*$ , dove  $i$  rappresenta l'unità statistica di primo livello (dottore di ricerca), con  $i=1, \dots, N_j$ , e  $j$  rappresenta l'unità di secondo livello (corso di dottorato), con  $j=1, \dots, C$  (capitolo 2, paragrafo 2.8.6). In particolare, le variabili latenti sottostanti sono definite a meno di una serie di parametri soglia, indicati con  $\tau_k$ :

$$y_{ij} = k \Leftrightarrow \tau_k < y_{ij}^* < \tau_{k+1},$$

dove  $k=0,1,\dots,K-1$  indica la modalità della variabile categorica osservata mentre i parametri soglia sono  $-\infty = \tau_0 < \tau_1 < \tau_2 < \dots < \tau_{K-1} < \tau_K = +\infty$ . Nel caso specifico, i parametri soglia sono nove poiché le variabili categoriche presentano ciascuna dieci modalità.

A differenza dei modelli tradizionali in cui si assume indipendenza tra tutte le unità osservate, in un modello di misura a due livelli l'indipendenza si assume soltanto per le unità di secondo livello (Muthén, 1991, 1994).

La formulazione a due livelli del modello di misura riferito al vettore di variabili risposta latenti continue  $\mathbf{y}_{ij}^*$  si può esprimere in forma ridotta nel modo seguente (capitolo 2, paragrafo 2.8.2):

$$\mathbf{y}_{ij}^* = \nu_b + \Lambda_b \eta_{bj} + \Lambda_{wij} \eta_{wij} + \boldsymbol{\varepsilon}_{bj} + \boldsymbol{\varepsilon}_{wij},$$

dove  $\nu_b$  rappresenta il vettore delle intercette,  $\eta_{bj}$  e  $\eta_{wij}$  sono le componenti

rispettivamente entro e tra i gruppi della variabile latente che si intende misurare,  $\Lambda_{bj}$  e  $\Lambda_{wij}$  sono i vettori dei coefficienti fattoriali che legano le variabili risposta osservate alle due componenti della variabile latente, mentre  $\epsilon_{bj}$  e  $\epsilon_{wij}$  sono i vettori degli errori di misura. Gli indici  $w$  e  $b$  stanno a significare rispettivamente “within” (riferito alle unità di primo livello) e “between” (riferito alle unità di secondo livello). Il modello assume che le componenti siano tra loro incorrelate,  $\text{cov}(\eta_{bj}, \eta_{wij}) = 0$  e che ciascun indicatore osservato dipenda da una sola variabile latente. Inoltre, ogni indicatore contiene un errore di misura che si assume incorrelato con le variabili latenti. Il modello di misura riferito alle variabili latenti che non prevedono una componente di gruppo (dovuta all’effetto del corso) si semplifica in quanto  $\nu_b = 0$ ,  $\eta_{bj} = 0$ ,  $\Lambda_{bj} = 0$  e  $\epsilon_{bj} = 0$ .

Il modello analizzato prevede le variabili latenti “Efficacia esterna” (E), “Soddisfazione” (S), “Aspettative” (A) e “Partecipazione lavorativa” (PL). Di queste, la variabile latente “Efficacia esterna” è misurata anche per le unità di secondo livello ed è indicata con (EB), “Efficacia esterna del corso”, mentre le variabili latenti “Soddisfazione”, “Aspettative” e “Partecipazione lavorativa” sono misurate solo per le unità di primo livello.

Il modello di misura riferito alla variabile latente “Efficacia esterna” “considera gli indicatori “Coerenza con gli studi fatti” (A305), “Utilizzo delle competenze acquisite” (A306) e “Rispondenza ai propri interessi culturali” (A308) e si può esprimere in forma estesa nel modo seguente:

$$y_{wA305ij}^* = \nu_{bA305} + \lambda_{bA305,EBj} \eta_{bEBj} + \lambda_{wA305,Eij} \eta_{wEij} + \epsilon_{bA305,EBj} + \epsilon_{wA305,Eij},$$

$$y_{wA306ij}^* = \nu_{bA306} + \lambda_{bA306,EBj} \eta_{bEBj} + \lambda_{wA306,Eij} \eta_{wEij} + \epsilon_{bA306,EBj} + \epsilon_{wA306,Eij},$$

$$y_{wA308ij}^* = \nu_{bA308} + \lambda_{bA308,EBj} \eta_{bEBj} + \lambda_{wA308,Eij} \eta_{wEij} + \epsilon_{bA308,EBj} + \epsilon_{wA308,Eij}.$$

Le equazioni precedenti rendono esplicita la scomposizione della variabile latente “Efficacia esterna” in una componente di variabilità dovuta al corso  $j$ , riassunta nei coefficienti  $\lambda_{bA...-EBj}$ , ed una componente di variabilità individuale, riassunta nei coefficienti  $\lambda_{wA...-EBj}$ .

Il modello di misura riferito alla variabile latente “Aspettative” considera gli indicatori “Prospettive di guadagno” (A302), “Possibilità di carriera” (A303) e “Stabilità/sicurezza del lavoro” (A304) e si può esprimere in forma estesa nel modo seguente:

$$y_{wA302ij}^* = \lambda_{wA302,Eij} \eta_{wEij} + \varepsilon_{wA302,Eij} ,$$

$$y_{wA303ij}^* = \lambda_{wA303,Eij} \eta_{wEij} + \varepsilon_{wA303,Eij} ,$$

$$y_{wA304ij}^* = \lambda_{wA304,Eij} \eta_{wEij} + \varepsilon_{wA304,Eij} .$$

Il modello di misura riferito alla variabile latente “Partecipazione lavorativa” considera gli indicatori “Acquisizione di professionalità” (A307), “Indipendenza e autonomia sul lavoro” (A309) e “Coinvolgimento nei processi decisionali” (A310) e si può esprimere in forma estesa nel modo seguente:

$$y_{wA307ij}^* = \lambda_{wA307,Eij} \eta_{wEij} + \varepsilon_{wA307,Eij} ,$$

$$y_{wA309ij}^* = \lambda_{wA309,Eij} \eta_{wEij} + \varepsilon_{wA309,Eij} ,$$

$$y_{wA310ij}^* = \lambda_{wA310,Eij} \eta_{wEij} + \varepsilon_{wA310,Eij} .$$

Il modello di misura riferito alla variabile latente “Soddisfazione” considera gli indicatori “Rispetto alle mansioni svolte” (A316), “Rispetto al titolo di dottore” (A317) e “E nel suo complesso” (A318) e si può esprimere in forma estesa nel modo seguente:

$$y_{wA316ij}^* = \lambda_{wA316-Sij} \eta_{wSij} + \varepsilon_{wA316-Sij} ,$$

$$y_{wA317ij}^* = \lambda_{wA317-Sij} \eta_{wSij} + \varepsilon_{wA317-Sij} ,$$

$$y_{wA318ij}^* = \lambda_{wA318-Sij} \eta_{wSij} + \varepsilon_{wA318-Sij} .$$

Le equazioni del modello di misura riferite alle variabili latenti “Soddisfazione” (S), “Aspettative” (A) e “Partecipazione lavorativa” (PL) non prevedono i coefficienti fattoriali dovuti alla componente di gruppo.

Il modello di misura può essere riassunto per tutte le variabili latenti, nel modo

seguinte:

$$\begin{bmatrix} y_{wA305ij}^* \\ y_{wA306ij}^* \\ y_{wA308ij}^* \\ y_{wA302ij}^* \\ y_{wA303ij}^* \\ y_{wA304ij}^* \\ y_{wA307ij}^* \\ y_{wA309ij}^* \\ y_{wA310ij}^* \\ y_{A316ij}^* \\ y_{A317ij}^* \\ y_{A318ij}^* \end{bmatrix}_{(12 \times 1)} = \begin{bmatrix} v_{bA305} \\ v_{bA306} \\ v_{bA307} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{(12 \times 1)} + \begin{bmatrix} \lambda_{bA305-EBj} & 0 & 0 & 0 \\ \lambda_{bA306-EBj} & 0 & 0 & 0 \\ \lambda_{bA307-EBj} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}_{(12 \times 4)} \begin{bmatrix} \eta_{bEBj} \\ \eta_{bABj} \\ \eta_{bPLBj} \\ \eta_{bSBj} \end{bmatrix}_{(4 \times 1)} + \begin{bmatrix} \lambda_{wA305-Eij} & 0 & 0 & 0 \\ \lambda_{wA306-Eij} & 0 & 0 & 0 \\ \lambda_{wA307-Aij} & 0 & 0 & 0 \\ 0 & \lambda_{wA302-Aij} & 0 & 0 \\ 0 & \lambda_{wA303-Aij} & 0 & 0 \\ 0 & \lambda_{wA304-Aij} & 0 & 0 \\ 0 & 0 & \lambda_{wA307-PLij} & 0 \\ 0 & 0 & \lambda_{wA308-PLij} & 0 \\ 0 & 0 & \lambda_{wA305-PLij} & 0 \\ 0 & 0 & 0 & \lambda_{wA305-Sij} \\ 0 & 0 & 0 & \lambda_{wA305-Sij} \\ 0 & 0 & 0 & \lambda_{wA305-Sij} \end{bmatrix}_{(12 \times 4)} \begin{bmatrix} \eta_{wEij} \\ \eta_{wAij} \\ \eta_{wPLij} \\ \eta_{wSij} \end{bmatrix}_{(4 \times 1)} + \begin{bmatrix} \varepsilon_{wA305ij} \\ \varepsilon_{wA306ij} \\ \varepsilon_{wA308ij} \\ \varepsilon_{wA302ij} \\ \varepsilon_{wA303ij} \\ \varepsilon_{wA304ij} \\ \varepsilon_{wA307ij} \\ \varepsilon_{wA309ij} \\ \varepsilon_{wA310ij} \\ \varepsilon_{A316ij} \\ \varepsilon_{A317ij} \\ \varepsilon_{wA318ij} \end{bmatrix}_{(12 \times 1)}.$$

### 3.5.2 Il modello a variabili latenti

Il modello a variabili latenti è un sistema di equazioni lineari in cui sono specificate le relazioni ipotizzate tra le variabili latenti e tra queste e le variabili esplicative osservate.

La struttura gerarchica a due livelli è rappresentata rispettivamente da un'equazione che formalizza le relazioni entro i gruppi e da un'equazione che formalizza le relazioni tra i gruppi, secondo quanto riportato di seguito (capitolo 2, paragrafo 2.8.6):

$$\eta_{wij} = \mathbf{B}_w \eta_{wij} + \mathbf{\Gamma}_w \mathbf{x}_{wij} + \zeta_{wij},$$

$$\eta_{bj} = \alpha_b + \mathbf{B}_b \eta_{bj} + \mathbf{\Gamma}_b \mathbf{x}_{bj} + \zeta_{bj}.$$

dove, rispettivamente per il modello entro ed il modello tra i gruppi,  $\eta_{wij}$  e  $\eta_{bj}$  rappresentano i vettori delle variabili latenti,  $\mathbf{B}_w$  e  $\mathbf{B}_b$  le matrici dei coefficienti che legano le variabili latenti tra di loro,  $\mathbf{x}_{wij}$  e  $\mathbf{x}_{bj}$  i vettori delle variabili esplicative osservate,  $\mathbf{\Gamma}_w$  e  $\mathbf{\Gamma}_b$  le matrici che legano le variabili esplicative alle variabili latenti mentre  $\zeta_{wij}$  e  $\zeta_{bj}$  sono i termini di errore.

Il modello a variabili latenti è specificato in base alle considerazioni teoriche espresse nei paragrafi 3.4.1 e 3.4.2. La variabile latente “Efficacia esterna” è l'unica che prevede una parte entro ed una tra i gruppi, “Efficacia esterna del corso”. Le relazioni

che le riguardano si possono scrivere in forma estesa rispettivamente nel modo seguente:

$$\eta_{wEij} = \beta_{wE,S} \eta_{Sij} + \beta_{E,A} \eta_{Aij} + \beta_{E,PL} \eta_{PL} + \gamma_{wE,M} x_{Mij} + \gamma_{wE,Etd} x_{Etdij} + \gamma_{wE,Z05} x_{Z05ij} + \gamma_{wE,Z06} x_{Z06ij} + \gamma_{wE,A301} x_{A301ij} + \zeta_{ij}$$

$$\eta_{EBj} = \gamma_{bEB,Unipriv} x_{Uniprivj} + \gamma_{bEB,Z04} x_{Z04j} + \gamma_{bEB,Area} x_{Areaj} + \zeta_j$$

La variabile latente “Soddisfazione” presenta il sistema di relazioni espresso in forma estesa nel modo seguente:

$$\eta_{Sij} = \beta_{wS,A} \eta_{Aij} + \beta_{wS,PL} \eta_{PLij} + \gamma_{wA301} x_{A301ij} + \gamma_{wM} x_{Mij} + \zeta_{ij}$$

La variabile latente “Aspettative” presenta il sistema di relazioni espresso in forma estesa nel modo seguente:

$$\eta_{Aij} = \gamma_{A301} x_{A301ij} + \zeta_{ij}.$$

### 3.6 I risultati

Il modello finale è stato stimato considerando 3053 unità di primo livello (dottori di ricerca) raggruppate in 1323 unità di secondo livello (corsi di dottorato). Il numero medio di dottori per ciascun corso è risultato pari a 2,31 mentre il numero mediano è stato pari a 2. Poiché i corsi di dottorato sono tipicamente composti da poche unità, alcuni di questi non sono rappresentati a causa delle mancate risposte.

Le variabili risposta (o dipendenti) osservate sono complessivamente dodici: “Coerenza con gli studi fatti (A305); “Utilizzo delle competenze acquisite” (A306); “Rispondenza ai propri interessi culturali” (A308); “Rispetto alle mansioni svolte” (A316); “Rispetto al titolo di dottore” (A317); “E nel suo complesso” (A318); “Prospettive di guadagno” (A302); “Possibilità di carriera” (A303); “Stabilità/sicurezza del lavoro” (A304); “Acquisizione di professionalità” (A307); “Indipendenza e autonomia sul lavoro” (A309); “Coinvolgimento nei processi decisionali” (A310). Tali variabili presentano ciascuna dieci modalità ordinate (dove 1 = per niente e 10 =

moltissimo) e sono legate alle variabili latenti da un modello di misura probit multinomiale (capitolo 2, paragrafo 2.3.1). Per ogni variabile risposta categorica sono stimati tanti parametri soglia per quante sono le modalità meno uno.

Le variabili latenti sono cinque e sono di tipo continuo: “Efficacia esterna” (E), “Soddisfazione” (S), “Aspettative” (A), “Partecipazione lavorativa” (PL) ed “Efficacia esterna del corso” (EB). Le variabili risposta osservate riferite alle variabili latenti “Efficacia esterna” (E) ed “Efficacia esterna del corso” (EB) sono le stesse.

Le variabili osservate esplicative (o indipendenti) sono tredici, di cui cinque relative alle unità di primo livello e otto relative alle unità di secondo livello. Le variabili esplicative di primo livello sono quattro di tipo binario ed una di tipo categorico ordinale: “Maschi” (M), (Femmine = 0 e Maschi = 1); “Età (fino a 32 anni)” (Eta), (Più di 32 anni = 0 e Fino a 32 anni = 1); “Periodo di studi all’estero” (Z05), (No = 0 e Sì = 1); “Attività lavorativa durante il dottorato” (Z06), (No = 0 e Sì = 1); “Livello di soddisfazione per retribuzione” (A301), (dieci modalità dove 1 = per niente e 10 = moltissimo).

Il modello ad equazioni strutturali a due livelli è stato stimato utilizzando il software *Mplus* 5.21, considerando lo stimatore ad informazione limitata indicato con l’acronimo “WLSMV” (capitolo 2, paragrafi 2.4.6 e 2.8.6), che si avvale della funzione di adattamento dei minimi quadrati ponderati e restituisce la statistica di adattamento robusta *mean and variance adjusted chi-square* basata sulla distribuzione chi-quadrato (capitolo 2, paragrafo 2.6.6).

I risultati della stima sono riportati nelle tabelle 3.1, 3.2, 3.3, 3.4 e 3.5. Queste comprendono le stime dei coefficienti non standardizzati (Stima), gli errori standard associati (Errore standard), il rapporto tra le stime e gli errori standard (Stima/ES), i p valori, le stime dei coefficienti standardizzati rispetto alle latenti (Std), le stime dei coefficienti standardizzati rispetto alle variabili y e x (StimaYX) ed i valori dell’R-quadrato. Il modello finale è stato inoltre rappresentato nella figura 3.5.

### 3.6.1 L’analisi della struttura a due livelli

Per ciascun indicatore della variabile latente “Efficacia esterna” è stato inizialmente calcolato il coefficiente di correlazione intraclasse che consente di scomporre la variabilità totale delle risposte osservate nei due livelli della gerarchia e di quantificare

la quota di variabilità totale dovuta al livello di gruppo (Snijders e Bosker, 1999). I valori ottenuti sono stati:

- “Coerenza con gli studi fatti” (A305)  $\rho_{A305} = 0,02$ ,
- “Utilizzo delle competenze acquisite” (A306)  $\rho_{A306} = 0,04$ ,
- “Rispondenza ai propri interessi culturali” (A308)  $\rho_{A308} = 0,01$ .

I valori osservati per ogni variabile indicano un grado di associazione piuttosto moderato nelle risposte per ciascun corso di dottorato.

La stima del modello nullo, in cui non sono presenti variabili esplicative, consente di calcolare anche il coefficiente di correlazione intraclasse per la variabile latente endogena “Efficacia esterna”. In questo modo è possibile suddividere la varianza totale della variabile latente in una componente entro i gruppi ed una componente tra i gruppi e determinare la quota di variabilità della variabile latente “Efficacia esterna” dovuta alle differenze tra i corsi rispetto alla variabilità totale. In particolare, riferendosi ad una generica variabile latente, la varianza totale si può scomporre nel modo seguente (capitolo 2, paragrafo 2.9.2):

$$Var(\eta_{ij}) = \psi_T = \psi_B + \psi_W,$$

dove  $Var(\eta_{Bj}) = \psi_B$  e  $Var(\eta_{Wij}) = \psi_W$ . Da questa relazione, il coefficiente di correlazione intraclasse (ICC) per la variabile latente “Efficacia esterna” si ottiene nel modo seguente:

$$ICC_E = \rho_E = \frac{\psi_B}{\psi_B + \psi_W} = \frac{0,619}{19,049} = 0,03.$$

Il valore ottenuto è stato di  $\rho_E = 0,03$  e, anche in questo caso, è indicativo di un livello moderato di associazione nelle risposte osservate per ciascun corso di dottorato. Tuttavia, è utile osservare che, rispetto al coefficiente di correlazione intraclasse calcolato per una variabile osservata, il coefficiente di correlazione intraclasse calcolato



per una variabile latente risulta attenuato a causa della presenza dell'errore di misura che va ad incrementare la componente di variabilità residua riassunta in  $\psi_W$  e quindi il denominatore del rapporto.

Il valore basso assunto dal coefficiente di correlazione intraclasse e la scarsa numerosità media dei gruppi, comportano un valore basso anche dell'effetto del disegno (capitolo 2, paragrafo 2.7.2), la cui grandezza è inferiore a due.

Per stabilire se la componente di varianza tra i gruppi è significativa, si sottopongono a verifica le seguenti ipotesi:

$$H_0 : \psi_B = 0 \quad \text{contro} \quad H_1 : \psi_B > 0 ,$$

dove l'ipotesi nulla equivale a stabilire l'assenza della componente di variabilità tra i gruppi mentre l'ipotesi alternativa è unidirezionale e stabilisce che questa componente è positiva perché una varianza può assumere solo valori positivi. Per questo motivo, il p valore osservato in corrispondenza del valore associato alla statistica test utilizzata dovrà essere dimezzato (Snijders e Bosker, 1999).

Per la verifica di ipotesi sul parametro  $\psi_B$  è stato utilizzato il test di devianza o test del rapporto di verosimiglianza. La devianza è definita come due volte il logaritmo della verosimiglianza e può essere considerata una misura della mancanza di adattamento tra il modello ed i dati osservati. Tale misura è utile per confrontare modelli diversi stimati sugli stessi dati. Per questo si stima prima il modello nullo  $M_0$  che non prevede la componente di variabilità tra i gruppi e poi il modello alternativo  $M_1$  che invece la prevede. Per ogni modello si ottengono i valori della verosimiglianza, indicata rispettivamente con  $L_0$  ed  $L_1$  e si pone a verifica l'ipotesi che il modello vero sia  $M_0$  con  $m_0$  gradi di libertà e quello alternativo  $M_1$ , con  $m_1$  gradi di libertà. La differenza tra le due devianze corrispondenti è una statistica test che segue una distribuzione chi-quadrato con  $m_0 - m_1$  gradi di libertà:

$$D = -2 \log \left( \frac{L_0}{L_1} \right) = -2(\log L_0 - \log L_1).$$

Il valore osservato della statistica test porta a rifiutare l'ipotesi nulla ( $p < 0.001$ ) il che significa che l'effetto dei corsi di dottorato sulla variabile latente endogena "Efficacia esterna" è significativo, nonostante il suo valore sia piuttosto basso.

Questo risultato, unito all'interesse per la possibilità di analizzare in un unico modello statistico variabili misurate a livelli diversi, riferite rispettivamente alle unità di primo e di secondo livello, giustifica l'analisi a due livelli anche se il grado di dipendenza delle risposte per ciascun corso di dottorato è piuttosto contenuto.

### 3.6.2 L'analisi del modello di misura

L'analisi dei valori dei coefficienti fattoriali che definiscono la relazione tra le variabili latenti e le variabili risposta osservate (modello di misura) mostra che la variabile latente "Efficacia esterna" è ben misurata dalle variabili utilizzate per ciascun livello della gerarchia (Tabella 3.1). Per quanto riguarda le stime non standardizzate (Stima), il primo coefficiente fattoriale di ogni variabile latente è fissato ad uno per problemi di identificazione.

Tabella 3.1. Modello di misura: stime dei coefficienti fattoriali.

	Stima	Errore standard	Stima/ES	P valore	Std	StdYX	R-quadrato
<b>Modello entro i gruppi (WITHIN)</b>							
<b>Efficacia esterna (E):</b>	-	0,025	28,023	0,000	-	-	0,696
Coerenza con gli studi fatti (A305)	1,000	-	-	-	1,687	0,860	0,740
Utilizzo delle competenze acquisite (A306)	2,836	0,543	5,226	0,000	4,784	0,979	0,958
Rispondenza ai propri interessi culturali (A308)	1,062	0,072	14,687	0,000	1,791	0,873	0,762
<b>Soddisfazione (S):</b>	-	0,023	33,332	0,000	-	-	0,780
Rispetto alle aspettative iniziali (A315)	1,000	-	-	-	1,418	0,817	0,668
Rispetto al titolo di dottore (A317)	1,250	0,091	13,697	0,000	1,773	0,871	0,759
E nel suo complesso (A318)	1,657	0,164	10,122	0,000	2,350	0,920	0,847
<b>Aspettative (A):</b>	-	0,016	26,735	0,000	-	-	0,415
Prospettive di guadagno (A302)	1,000	-	-	-	3,045	0,950	0,903
Possibilità di carriera (A303)	0,637	0,063	10,084	0,000	1,940	0,889	0,790
Stabilità/sicurezza del lavoro (A304)	0,280	0,025	11,166	0,000	0,852	0,648	0,421
<b>Partecipazione lavorativa (PL):</b>	-	-	-	-	-	-	-
Acquisizione di professionalità (A307)	1,000	-	-	-	2,250	0,914	0,835
Indipendenza e autonomia sul lavoro (A309)	0,527	0,057	9,226	0,000	1,185	0,764	0,584
Coinvolgimento nei processi decisionali (A310)	0,444	0,044	10,123	0,000	0,999	0,707	0,499
<b>Modello tra i gruppi (BETWEEN)</b>							
<b>Efficacia esterna (EB):</b>	-	0,304	3,283	0,001	-	-	0,998
Coerenza con gli studi fatti (A305)	1,000	-	-	-	0,402	0,993	0,986
Utilizzo delle competenze acquisite (A306)	2,805	0,795	3,531	0,000	1,127	0,995	0,991
Rispondenza ai propri interessi culturali (A308)	0,267	0,355	0,753	0,004	0,107	0,962	0,926

Entro i gruppi, i coefficienti standardizzati rispetto alle variabili  $y$  ed  $x$  ( $\text{StdYX}$ )<sup>109</sup> riferiti agli indicatori osservati variano tra 0,860, per la “Coerenza con gli studi fatti” (A305) e 0,979 per l’“Utilizzo delle competenze acquisite” (A306). I valori dell’R-quadrato associati sono di conseguenza piuttosto elevati ed assumono valori compresi tra 0,740 per la “Coerenza con gli studi fatti” (A305) e 0,958 per l’“Utilizzo delle competenze acquisite” (A306).

Tra i gruppi, i valori dei coefficienti fattoriali sono anche più elevati rispetto a quelli osservati per la parte entro i gruppi e sono superiori a 0,96 per tutti gli indicatori osservati. Di conseguenza, i valori dell’R-quadrato sono tutti maggiori di 0,92.

Il modello di misura sembra adeguato anche per le altre variabili latenti, anche se per queste i valori dei coefficienti fattoriali sono leggermente più bassi rispetto a quelli registrati per la variabile latente endogena.

La “Soddisfazione” presenta valori dei coefficienti fattoriali che variano tra 0,817 dell’indicatore “Rispetto alle aspettative iniziali” (A317) e 0,889 dell’indicatore “E nel suo complesso” (A318) con valori dell’R-quadrato rispettivamente di 0,668 e 0,847.

Le “Aspettative” hanno invece coefficienti fattoriali più bassi e variano tra 0,648 dell’indicatore “Stabilità/sicurezza del lavoro” (A304) e 0,950 dell’indicatore “Prospettive di guadagno” (A302) con valori dell’R-quadrato rispettivamente di 0,421 e 0,903. I valori piuttosto bassi osservati per la “Stabilità/sicurezza del lavoro” indicano che la quota di variabilità non spiegata dalla variabile latente “Aspettative” è piuttosto elevata.

La “Partecipazione lavorativa” presenta invece coefficienti fattoriali che variano tra 0,707 dell’indicatore “Coinvolgimento nei processi decisionali” (A310) e 0,914 dell’indicatore l’“Acquisizione di professionalità” (A307) con valori dell’R-quadrato rispettivamente di 0,499 e 0,835. Anche per l’indicatore “Coinvolgimento nei processi decisionali”, si deve osservare che la quota di variabilità non spiegata dalla variabile latente è piuttosto elevata.

---

<sup>109</sup> Per gli indicatori osservati che si assumono dipendenti da una sola variabile latente, i coefficienti fattoriali standardizzati rappresentano le correlazioni tra le variabili osservate e le variabili latenti ed indicano la quota di variabilità di una variabile osservata dovuta alla presenza della variabile latente.

### 3.6.3 L'analisi delle relazioni di primo livello

L'analisi delle relazioni tra le variabili latenti (Tabella 3.2) riferite alle unità di primo livello (dottori di ricerca), mostra che la “Soddisfazione” ( $\beta = +0,842; p < 0.001$ ) ha un effetto diretto positivo significativo sulla variabile latente endogena “Efficacia esterna”, a conferma dell'ipotesi formulata. Anche la “Partecipazione lavorativa” esercita un effetto diretto sull’“Efficacia esterna” ( $\beta = 0,175; p = 0,004$ ) mentre le “Aspettative” hanno un effetto diretto significativo di segno negativo ( $\beta = -0,366; p < 0,001$ ).

Tabella 3.2. Modello entro i gruppi (within): stime dei coefficienti.

	Stima	Errore standard	Stima/ES	P valore	Std	StdYX
<b>Regressioni tra le variabili latenti</b>						
<b>Efficacia esterna (E) rispetto a:</b>	-	-	-	-	-	-
Soddisfazione (S)	1,001	0,107	9,343	0,000	0,842	0,842
Aspettative (A)	-0,203	0,039	-5,145	0,000	-0,366	-0,366
Partecipazione lavorativa (PL)	0,131	0,046	2,861	0,004	0,175	0,175
<b>Soddisfazione (S) rispetto a:</b>	-	-	-	-	-	-
Aspettative (A)	0,240	0,025	9,533	0,000	0,515	0,515
Partecipazione lavorativa (PL)	0,402	0,045	8,850	0,000	0,638	0,638
<b>Regressioni tra le variabili latenti e le variabili esplicative osservate</b>						
<b>Efficacia esterna (E) rispetto a:</b>	-	-	-	-	-	-
Maschi (M)	-0,121	0,055	-2,216	0,027	-0,072	-0,036
Età (fino a 32 anni) (Eta)	0,054	0,067	0,799	0,424	0,032	0,016
Periodo di studi all'estero (Z05)	0,241	0,069	3,503	0,000	0,143	0,070
Attività lavorativa durante il dottorato (Z06)	-0,044	0,067	-0,640	0,522	-0,026	-0,016
<b>Aspettative (A) rispetto a:</b>	-	-	-	-	-	-
Livello di soddisfazione per retribuzione (A301)	2,419	0,186	12,995	0,000	0,794	0,648
<b>Soddisfazione (S) rispetto a:</b>	-	-	-	-	-	-
Livello di soddisfazione per retribuzione (A301)	0,232	0,042	5,504	0,000	0,164	0,124
Maschi (M)	0,082	0,053	1,538	0,124	0,058	0,019

Le “Aspettative” ( $\beta = +0,515; p < 0,001$ ) e la “Partecipazione lavorativa” ( $\beta = +0,638; p < 0,001$ ) hanno anche un effetto diretto significativo di segno positivo sulla “Soddisfazione”.

In conseguenza di queste relazioni, le “Aspettative” e la “Partecipazione lavorativa” esercitano anche un effetto indiretto sull’“Efficacia esterna” attraverso la “Soddisfazione”. Per quanto riguarda le “Aspettative”, tale effetto è dato dal prodotto tra gli effetti diretti delle “Aspettative” sulla “Soddisfazione” e della “Soddisfazione” sull’“Efficacia esterna” ( $0,515 \times 0,842 = 0,434$ ). Analogamente, si può calcolare anche l'effetto indiretto della “Partecipazione lavorativa” sull’“Efficacia esterna”, dato dal prodotto degli effetti diretti della “Partecipazione lavorativa” sulla “Soddisfazione”

(0,638) e della “Soddisfazione” sull’”Efficacia esterna” ( $0,638 \times 0,842 = 0,537$ ).

L’effetto totale delle “Aspettative” sull’”Efficacia esterna” è dato dalla somma tra gli effetti diretti e gli effetti indiretti ed assume un valore piuttosto contenuto ( $-0,366 + 0,434 = 0,068$ ). L’effetto totale della “Partecipazione lavorativa” sull’”Efficacia esterna” assume invece un valore piuttosto elevato ( $0,175 + 0,537 = 0,712$ ).

Per quanto riguarda le variabili esogene osservate, si registra l’effetto di genere significativo nei confronti della variabile latente endogena “Efficacia esterna” ( $\beta = -0,036; p = 0,027$ ) a sfavore dei maschi rispetto alle femmine. Non si registra invece un effetto significativo della variabile “Età” ( $\beta = -0,016; p = 0,424$ ).

Aver svolto un “Periodo di studi all’estero” (Z05) ha un effetto positivo significativo sull’”Efficacia esterna” ( $\beta = +0,070; p < 0,001$ ). Invece, aver svolto un’”Attività lavorativa durante il dottorato” (Z06) non produce effetti significativi ( $\beta = -0,016; p = 0,522$ ).

Per quanto riguarda le altre variabili latenti, si registra l’effetto diretto positivo della variabile “Grado di soddisfazione per il livello di retribuzione attuale” (A301) nei confronti sia delle “Aspettative” ( $\beta = +0,648; p < 0,001$ ) e che della “Soddisfazione” ( $\beta = +0,124; p < 0,001$ ).

#### 3.6.4 *L’analisi delle relazioni di secondo livello*

L’analisi degli effetti esercitati dalle variabili esplicative della variabile latente “Efficacia esterna del corso” (Tabella 3.3) evidenzia il valore significativo positivo del coefficiente relativo alla variabile “% di chi si iscriverebbe allo stesso corso” (Z04) ( $\beta = +0,894; p < 0,001$ ). Tale variabile, pur con le dovute cautele (paragrafo 3.4.2), può considerarsi indicativa della “qualità” del corso di dottorato frequentato ed evidenzia che i corsi maggiormente efficaci sono quelli in cui i dottori si iscriverebbero di nuovo.

Non si osserva un effetto significativo dei corsi organizzati presso università private rispetto ai corsi organizzati in università pubbliche ( $\beta = +0,150; p = 0,168$ ).

Per quanto riguarda le aree disciplinari, si evidenzia un effetto significativo negativo sull’”Efficacia esterna” dei corsi appartenenti all’Area 2 (Scienze chimiche e della terra) rispetto ai corsi dell’Area 1 (Scienze matematiche, informatiche e fisiche)

$(\beta = -0,291; p = 0,022)$ .

*Tabella 3.3. Modello tra i gruppi: le variabili esplicative dell'”Efficacia esterna dei corsi”.*

	Stima	Errore standard	Stima/ES	P valore	Std	StdYX
<b>Efficacia esterna (EB) rispetto a:</b>	-	-	-	-	-	-
Università privata (Unipriv)	0,223	0,162	1,378	0,168	0,555	0,155
% di chi si iscriverebbe allo stesso corso (Z04)	0,009	0,002	5,596	0,000	0,022	0,894
Area 2	-0,415	0,182	-2,287	0,022	-1,034	-0,291
Area 3	-0,220	0,149	-1,483	0,138	-0,548	-0,256
Area 4	-0,130	0,155	-0,843	0,399	-0,325	-0,119
Area 5	-0,291	0,159	-1,829	0,067	-0,724	-0,285
Area 6	0,207	0,214	0,966	0,334	0,515	0,134
Area 7	-0,186	0,179	-1,041	0,298	-0,463	-0,145

Per quanto riguarda le altre aree disciplinari, si deve osservare un livello di “Efficacia esterna” inferiore rispetto all’Area 1 anche per i corsi dell’Area 5 (Scienze dell’antichità, filologico-letterarie e storico-artistiche, Scienze storiche, filosofiche, pedagogiche e psicologiche), anche se il livello di significatività è di poco superiore al cinque per cento ( $\beta = -0,285; p = 0,067$ ).

### 3.6.5 Gli indici di adattamento del modello

I valori ottenuti dagli indici di adattamento del modello sono, nel complesso, abbastanza incoraggianti (Tabella 3.4 e Tabella 3.5).

La statistica Chi-quadrato del modello presenta un valore significativo ( $\chi^2 = 294,005; df = 29; p < 0,001$ ) ed indica che la matrice di covarianza implicata dal modello,  $\Sigma(\hat{\theta})$ , è diversa dalla matrice di covarianza campionaria, **S** (Tabella 3.4). Tuttavia, la sensibilità di questa statistica alla numerosità campionaria tende a far emergere valori della statistica test significativi anche quando le differenze tra le matrici sono lievi (capitolo 2 paragrafo 2.6.1).

Per valutare l’adeguatezza del modello si è fatto quindi ricorso alle misure di adattamento approssimativo (Tabella 3.5). Le misure considerate sono state l’indice SRMR (capitolo 2 paragrafo 2.6.2), che si basa sulla differenza tra la matrice di covarianza campionaria e quella implicata, l’indice RMSA che può intendersi come una misura di parsimonia corretta e due misure di adattamento comparativo (o

incrementale), gli indici CFI (*Comparative Fit Index*) e TLI o NNFI (*Non Normed Fit Index*) (vedi capitolo 2 paragrafo 2.6.3). Tutti questi indici rappresentano misure descrittive continue di corrispondenza tra il modello ed i dati osservati.

*Tabella 3.4. Statistica test del modello*

Indice	Valore	Gradi di libertà	P valore
Chi-quadrato modello stimato	294,005	29	0,000
Chi-quadrato modello di base	6579,450	36	0,000

Gli indici CFI e TLI misurano il miglioramento relativo dell'adattamento di un modello in confronto a quello di un modello di base. Il modello di base è di solito quello di indipendenza ed assume una matrice di covarianza nulla, variabili osservate misurate senza errore (cioè con varianze degli errori fissate a zero) e coefficienti fattoriali pari ad uno. In generale, tali indici sono tra quelli meno influenzati dalla numerosità campionaria (vedi capitolo 2 paragrafo 2.6.3). In particolare, l'indice TLI tende a penalizzare i modelli complessi e a premiare quelli parsimoniosi. I valori ottenuti sono stati rispettivamente di 0,960 per l'indice CFI e di 0,950 per l'indice TLI e sono entrambi indicativi di un adattamento complessivo accettabile che comporta un miglioramento importante del modello stimato rispetto a quello di indipendenza.

*Tabella 3.4. Indici alternativi di adattamento*

Indice	Valore
CFI (Comparative Fit Index)	0,960
TLI (Non Normed Fit Index)	0,950
RMSA (Root Mean Square Error of Approximation)	0,055
SRMR (Standardized Root Mean Square Residual)	
Entro i gruppi	0,114
Tra i gruppi	0,743

L'indice di adattamento RMSEA (*Root Mean Square Error of Approximation*) tende a decrescere all'aumentare del numero di gradi di libertà o all'aumentare della numerosità campionaria. Il valore osservato è stato di 0,055 e può considerarsi adeguato.

L'indice SRMR (*Standardized Root Mean Square Residual*) si basa sui residui standardizzati tra la matrice di covarianza osservata e quella implicata dal modello. Il suo valore è quasi accettabile per la componente entro i gruppi del modello (0,114) mentre non è soddisfacente per la componente tra i gruppi (0,743).

### 3.6.6 La rappresentazione grafica del modello finale

La rappresentazione grafica del modello finale (Figura 3.5) è stata realizzata riportando solo i coefficienti di regressione che indicano le relazioni tra le variabili (latenti ed osservate), tralasciando i coefficienti che legano le variabili latenti alle variabili osservate (coefficienti fattoriali) per rendere più chiara la comprensione dei risultati.

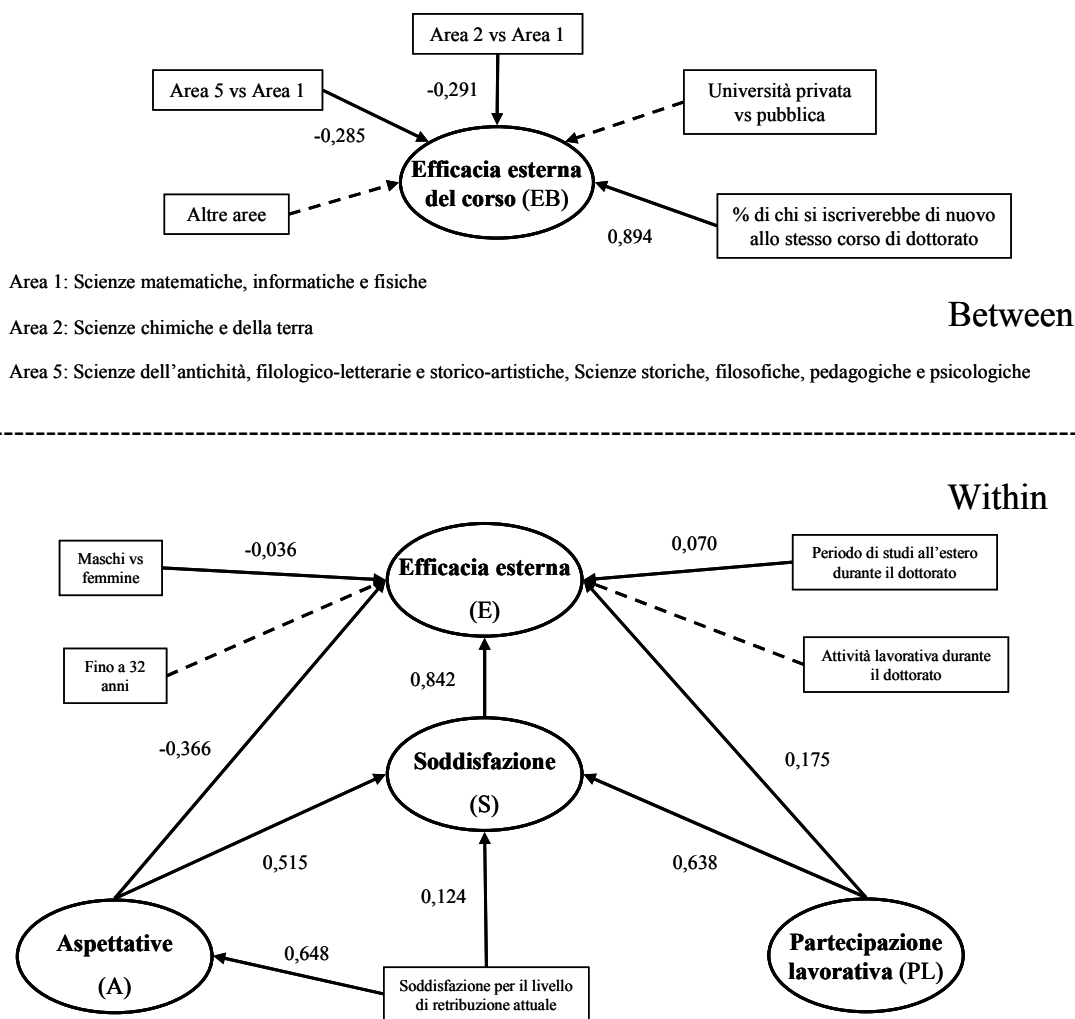


Figura 3.5. Rappresentazione grafica del modello finale.

I valori dei coefficienti di regressione sono riportati solo per le relazioni significative, per le quali sono associate le frecce direzionali continue. Per le relazioni che non sono risultate significative le frecce sono invece tratteggiate ed i valori dei coefficienti non sono stati riportati.



### **3.7 Limiti del modello proposto**

Il modello descritto consente di analizzare un sistema complesso di relazioni in cui sono presenti variabili latenti e variabili osservate. Dall'analisi dei risultati emergono considerazioni interessanti che possono offrire ulteriori spunti di riflessione. Tuttavia, il modello proposto soffre anche di alcuni limiti. Taluni di questi sono imputabili al metodo di analisi utilizzato, altri riguardano i dati a disposizione.

Per quanto riguarda il metodo, abbiamo già osservato che questo si basa sull'approccio tradizionale per estendere i modelli ad equazioni strutturali al caso multilivello e consente di stimare solo modelli a due livelli. Questo significa che non è possibile distinguere, se esiste, la parte di variabilità nelle risposte attribuibile ai livelli di raggruppamento gerarchicamente superiori al secondo (ateneo di appartenenza ed eventuali altri). Inoltre, l'approccio utilizzato è limitato ai modelli ad intercetta casuale, rimane quindi esclusa la possibilità di avere coefficienti delle variabili esplicative diversi (casuali) per ciascuna unità di secondo livello (corso di dottorato), una volta stabilito che questi differiscono tra loro. Ancora, non si possono stimare gli effetti di variabili di secondo livello (o variabili di contesto) sulle risposte di primo livello. Infine, non è possibile calcolare i valori delle intercette casuali associate a ciascun corso di dottorato. Questo valore, se stimato, si potrebbe intendere come una misura netta di performance rispetto all'”Efficacia esterna” e consentirebbe di stabilire una graduatoria dei corsi di dottorato.

Per quanto riguarda i dati a disposizione, dobbiamo innanzitutto osservare la scarsa numerosità delle unità di secondo livello (corsi di dottorato). Questo è dovuto principalmente al fatto che ad un corso di dottorato si accede per concorso e i posti a disposizione (con o senza borsa di studio) sono di solito limitati. A questo si deve aggiungere la presenza di mancate risposte che, nonostante siano numericamente limitate, finiscono per accentuare il problema. La conseguenza è che le stime dei parametri relative alla parte tra i gruppi del modello sono meno accurate. Infine, si deve considerare anche la scarsa disponibilità di variabili associate ai corsi di dottorato. Questo purtroppo riduce la capacità esplicativa del secondo livello del modello.

## ***Conclusioni***

L'obiettivo di questo lavoro di tesi è quello di costruire un modello ad equazioni strutturali multilivello con variabili latenti per valutare l'efficacia esterna del titolo di dottorato di ricerca. Per questo si è proceduto inizialmente ad una disamina molto sommaria della problematica riguardante la valutazione dei processi formativi offerti dal sistema universitario italiano ed alla descrizione dello strumento di misura proposto per l'analisi dell'efficacia esterna. In particolare, la misura utilizzata è una variabile latente ottenuta valutando il livello di soddisfazione per la condizione lavorativa attuale dei dottori di ricerca che hanno ottenuto il titolo nel 2008 rispetto agli indicatori "coerenza con gli studi fatti", "utilizzo delle competenze acquisite" e "rispondenza ai propri interessi culturali". Successivamente, per consentire una scelta critica e consapevole del modello da utilizzare, si è ritenuto di dare ampio spazio agli aspetti metodologici riguardanti i modelli ad equazioni strutturali. In particolare, si è soffermata l'attenzione sugli aspetti di specificazione, stima e adattamento del modello ai dati. Dopo aver descritto l'analisi dei modelli ad equazioni strutturali tradizionali ad un livello ed aver illustrato le caratteristiche generali dei modelli multilivello, si è proceduto con l'analisi dei modelli a due livelli e dei modelli ad equazioni strutturali generalizzati multilivello. Tuttavia, vi è piena consapevolezza che la parte metodologica, per quanto ampia, non possa ritenersi esaustiva. In particolare, non è stato affrontato il problema dell'identificazione del modello, per la trattazione del quale sarebbe stato necessario dedicare una parte specifica del lavoro di tesi, con il rischio di ampliare eccessivamente la parte metodologica. Inoltre, non è stata considerata l'analisi bayesiana dei modelli ad equazioni strutturali multilivello (Jedidi e Ansari, 2001).

L'analisi è stata realizzata utilizzando i dati dell'indagine "Condizione attuale e prospettive occupazionali dei dottori di ricerca", commissionata dal CNVSU al Dipartimento di Statistica "G. Parenti" dell'Università degli Studi di Firenze.

Il modello consente di rappresentare in maniera sintetica il sistema complesso di relazioni tra le variabili considerate e di spiegare l'"Efficacia esterna" tenendo conto della struttura di variabilità complessa presente nei dati. Per quanto riguarda le relazioni di primo livello (dottori di ricerca), l'analisi dei risultati evidenzia il ruolo decisivo esercitato dalla "Soddisfazione" lavorativa nei confronti dell'"Efficacia esterna". Questo significa che i dottori di ricerca più soddisfatti per la condizione lavorativa

attuale “rispetto alle aspettative iniziali”, “rispetto al titolo di dottorato di ricerca” e “nel suo complesso” sono coloro che presentano valori più elevati di “Efficacia esterna”, per i quali aumenta la possibilità che l’attività sia coerente con gli studi svolti” che le “competenze acquisite siano effettivamente utilizzate” e che l’attività lavorativa sia in linea con gli interessi culturali”. Sulla “Soddisfazione” agiscono direttamente le “Aspettative”, la “Partecipazione lavorativa” ed il “Grado di soddisfazione rispetto al livello di retribuzione attuale”. Tali variabili agiscono indirettamente anche sull’Efficacia esterna” in conseguenza della relazione diretta tra “Soddisfazione” ed “Efficacia esterna”. Per quanto riguarda le relazioni di secondo livello (corsi di dottorato di ricerca), l’analisi evidenzia che i corsi giudicati più “efficaci” sono quelli in cui i dottori sarebbero disposti ad iscriversi di nuovo mentre non si osserva alcuna differenza significativa tra i corsi organizzati dalle università private rispetto ai corsi organizzati dalle università pubbliche. Invece, per quanto riguarda l’area disciplinare, i corsi appartenenti all’area delle “Scienze chimiche e della terra” mostrano un livello di “Efficacia esterna” significativamente inferiore rispetto a quello dei corsi appartenenti all’area di riferimento, le “Scienze matematiche, fisiche ed informatiche”. Il livello di Efficacia esterna” è inferiore rispetto all’area di riferimento anche per i corsi dell’area delle “Scienze dell’antichità, filologico-letterarie e storico-artistiche, scienze storiche, filosofiche, pedagogiche e psicologiche”, anche se, in questo caso, il livello di significatività è di poco superiore al cinque per cento.

Il modello proposto presenta alcuni limiti (paragrafo 3.6.7). Quelli legati all’approccio utilizzato si potrebbero rimuovere ricorrendo ai modelli ad equazioni strutturali generalizzati multilivello, basati sulla stima di massima verosimiglianza (capitolo 2, paragrafo 2.9). Questo metodo è però impraticabile quando le variabili risposta osservate sono di natura categorica e gli effetti casuali sono numerosi, come nel caso in esame. Il modello utilizzato, basato sulla stima ad informazione limitata e sulla funzione di adattamento dei minimi quadrati ponderati è stato scelto proprio per tenere conto della natura categorica dei dati e per il proposito di stimare un modello piuttosto complesso. Una soluzione alternativa a questo problema potrebbe essere quella di trattare le variabili risposta osservate come continue, considerato che il livello di approssimazione che si ottiene migliora all’aumentare del numero di modalità assunte dalle variabili categoriche e diventa accettabile quando queste sono più di quattro

(capitolo 2, paragrafo 2.4.4). Gli effetti di questa possibilità sono stati considerati ma alla fine si è preferito privilegiare il rigore formale e rispettare la tipologia dei dati a disposizione. Tuttavia, un possibile sviluppo di questo lavoro potrebbe essere proprio quello di stimare un modello ad equazioni strutturali generalizzato seguendo l'ipotesi precedente. I limiti legati ai dati a disposizione sono invece irrisolvibili. La scarsa numerosità delle unità di secondo livello (corsi di dottorati) e la limitata disponibilità di caratteristiche che siano in grado di qualificarle rappresentano infatti i vincoli dell'analisi.

Infine, occorre ricordare che i risultati ottenuti dipendono inevitabilmente dalla misura di efficacia esterna che è stata utilizzata, in particolare dagli indicatori che sono stati scelti per rappresentarla. La misura proposta deve quindi intendersi come una misura relativa e non assoluta di efficacia. Pur con questa considerazione, vi è convinzione che i risultati ottenuti si possano rivelare di qualche utilità nell'attività di programmazione e gestione della formazione di terzo livello (dottorato) in Italia.

# Bibliografia

- Alker, H.R. (1969). *A typology of fallacies*. In M. Dogan & S. Rokkan (Eds.), Quantitative ecological analysis in the social sciences. Cambridge, Ma.: M.I.T. Press.
- Arbuckle, J.L. (1996). Full information estimation in the presence of incomplete data. In G.A. MarMarcoulides e R.E. Schumacker (Eds), *Advanced structural equation modelling: Issues and techniques* (pp. 243-277). Mahwah, NJ: Lawrence Erlbaum.
- Asparouhov, T. e Muthén, B. (2003). *Full-information maximum-likelihood estimation of general two-level latent variable models with missing data*. Mplus working paper.
- Asparouhov, T. e Muthén, B. (2006). *Multilevel Modeling of Complex Survey Data*. In Proceedings of the Joint Statistical Meeting in Seattle, August 2006. ASA section on Survey Research Methods, 2718-2726.
- Asparouhov, T. e Muthén, B. (2007). *Computationally efficient estimation of multilevel high-dimensional latent variable models*. In Proceedings of the Joint Statistical Meeting in Salt Lake City. ASA Section on Biometrics.
- Asparouhov, T. e Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397-438.
- Babakus, E. M., Ferguson, C.E & Jöreskog, K.G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, 24, 2228.
- Bartholomew, D.J. e Knott, M. (1999). *Latent variable models and factor analysis. Second edition*. London: Arnold.
- Bentler, P. M. & Dudgeon, P. (1996). Covariance Structure Analysis: Statistical Practice, Theory, and Directions. *Annual Review of Psychology*, 47, 563-592.
- Bentler, P.M. e Liang, J.J. (2003). Two-level mean and covariance structures:

- Maximum likelihood via an EM algorithm. In S. Reise e N. Duan (Eds), *Multilevel modelling: Methodological advances, issues, and applications* (pp. 53-70). Mahwah, NJ: Lawrence Erlbaum.
- Bollen, K.A. (1987). Total, direct, and indirect effects in structural equation models- *Sociological Methodology*, 17, 37-69.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K.A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605-634.
- Bollen, K.A. e Long, J.S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.
- Bollen, K.A., D.J. Bauer, S.L. Christ & M.C. Edwards (2010). An Overview of Structural Equations Models and Recent Extensions. In S. Kolenikov, D. Steinley, & L. Thombs. *Recent Developments in Social Science Statistics* (pp. 37-80). NY: Wiley.
- Browne, M.W. (1974). Generalized least-squares estimators in the analysis of covariance structures. *South African Statistical Journal*, 8, 1-24.
- Browne, M.W. (1982). Covariance structures. In D.M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). London: Cambridge University Press.
- Browne, M.W. (1984). Asymptotic distribution free methods in analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.
- Chiandotto, B. (2004). Sulla misura della qualità della formazione universitaria. *Studi e Note di Economia*. vol. 3.
- Chiandotto, B. (2008). Formazione universitaria e “valore” del capitale umano: un’analisi multilivello. In A. Cammelli, e G. Vittadini. *Capitale umano: esiti dell’istruzione universitaria*. Il Mulino, Bologna.
- Chou, C. e Bentler, P.M (1995). Estimates and test in structural equation modelling. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 37-55). Thousand Oaks, CA: Sage.
- Christofferson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40, 5-32.
- Cochran, W.G. (1977). *Sampling techniques*. 3<sup>rd</sup> ed. New York: Wiley.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to non-

- normality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16-29.
- Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.
- du Toit, H.C. & du Toit, M. (2008). Multilevel structural equation modeling. In J. de Leeuw & E. Meijer (Ed.), *Handbook of Multilevel Analysis* (pp. 435-479). New York, NY: Springer.
- Finch, J.F., West, S.G. & MacKinnon, D.P. (1997). Effects of sample size and non-normality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, 4, 87-107.
- Finney, S.J. & Di Stefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. Hancock & R.O. Mueller (Ed.), *Structural equation modelling: A second course* (pp. 345-383). Greenwich, CT: Information Age Publishing.
- Flora, D.B. e Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods*, 9, 466-491.
- Goldberg, A.S. e Duncan, O.D. (1973). *Structural equation models in the social sciences*. New York: Academic Press.
- Goldstein H. (2003). *Multilevel statistical models (Third Edition)*. London: Arnold.
- Goldstein, H.I. e McDonald, R.P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, 53, 455-467.
- Heck, R.H. & Thomas, S.L. (2009). *An introduction to multilevel modelling techniques (2<sup>nd</sup> ed.)*. New York, NY: Routledge.
- Hox, J.J. (2010). *Multilevel Analysis: Techniques and Applications. Second edition*. New York, NY: Routledge.
- Hox, J.J., Maas, C.G.M. & Brinkhuis, M.J.S. (2010). Test effect of estimation method and sample size in multilevel structural equation modelling. *Statistica Neerlandica*, 64, 157-170.
- Hu, L.T. e Bentler, P.M. (1995). Cutoff criteria for fit indexes in covariance structure analysis. *Psychometrika*, 47, 209-222.

- Hu, L.T. e Bentler, P.M. (1999). Evaluating model fit. In R.H. Hoyle (Ed), *Structural equation modelling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage.
- Hutchinson, S.R. & Olmos, A. (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling: A Multidisciplinary Journal*, 5, 344-364.
- Jöreskog, K.J. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.
- Jöreskog, K.J. (1973). A general method for estimating a linear structural equation. In A. Goldberg e O. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York: Academic Press.
- Jöreskog, K.J. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P.A. Krishnaiah (Ed), *Applications of statistics* (pp. 265-287). Amsterdam: North-Holland.
- Jöreskog, K.J. e Moustaki, I. (2001). Factor analysis of ordinal variables: a comparison of three approaches. *Multivariate Behavioral Research*, 36, 347-387.
- Kaplan, D. (2009). *Structural equation modelling (2<sup>nd</sup> ed.)*. Thousand Oakes, CA: Sage.
- Kaplan, D. & Elliott, P.R. (1997). A didactic example of multilevel structural equation modelling applicable to the study of organizations. *Structural Equation Modeling*, 4, 1-24.
- Kaplan, D. Kim, J.-S. e Kim, S.-Y. (2009). Multilevel Latent Variable Modeling: Current research and recent developments. In R. Millsap e A. Maydeus-Oliveras (Eds.), *Sage handbook of quantitative methods in psychology*. Thousand Oakes, CA: Sage.
- Kline, R.B. (2010). *Principles and practice of structural equation modelling (3<sup>rd</sup> ed.)*. New York, NY: The Guilford Press.
- Lee, S.-Y. (1990). Multilevel analysis of structural equation models. *Biometrika*, 77, 763-772.
- Lee, S.-Y. e Poon W.Y. (1998). Analysis of two-level structural equation models via EM type algorithms. *Statistica Sinica*, 8, 749-766.
- Lee, S.-Y. e Shi, J.Q. (2001). Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data. *Biometrics*, 57, 787-794.



- Lee, S.-Y. e Song, X.Y. (2004). Maximum likelihood analysis of a general latent variable model with hierarchically mixed data. *Biometrics*, 60, 624-636.
- Liang, J.J. e Bentler, P.M. (2004). An EM algorithm for fitting two-level structural equation models. *Psychometrika*, 69, 101-122.
- Linda, N., Lee, S.-Y. & Poon, W.-Y. (1993). Covariance structure analysis with three level data. *Computational Statistics & Data Analysis*, 15, 159-178.
- Lockeed, M.E. & Hanushek E.R. (1994). Concepts of Educational Efficiency and Effectiveness, in International Encyclopedia of Education, Second Edition.
- Longford, N.T. (1993). *Random coefficient models*. Oxford: Clarendon Press.
- Mardia, K. V. (1970) Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519-530.
- McDonald, R.P. e Goldstein, H.I. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Jpurnal of Mathematical and Statistical Psychology*, 42, 215-232.
- Metha, P.D. e Neale, M.C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, 10, 259-284.
- Mulaik, S.A. (2009). *Linear causal modelling with structural equations*. New York, NY: CRC.
- Muthén, B.O. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551-560.
- Muthén, B. (1983). Latent variable structural equation modelling with categorical data. *Journal of Econometrics*, 22, 43-65.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika*, 49, 115-132.
- Muthén, B. (1989). Latent variable modelling in heterogeneous populations. *Psychometrika*, 54, 557-585.
- Muthén, B. (1990). *Mean and covariance structure analysis of hierarchical data* (UCLA Statistics Series No. 62). Los Angeles: University of California.
- Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28, 338-354.
- Muthén, B. (1993). Goodness of fit with categorical and other non-normal variables. In K.A. Bollen e J.S. Long (Ed.), *Testing structural equation models* (pp. 205-243).

- Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research*, 22, 376-398.
- Muthén, B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological methodology 1997* (pp. 453-480). Washington, DC: American Sociological Association.
- Muthén, B. (2001). Latent variable mixture modeling. In G.A. Marcoulides & R. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp.1-33).
- Muthén, B., du Toit, S.H.C., Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modelling with categorical and continuous outcomes*. Unpublished manuscript.
- Muthén, B. e Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171-189.
- Muthén, B. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45, 19-30.
- Muthén, L. & Muthén, B. (1998-2010). *Mplus user's guide: Statistical analysis with Latent variables* (6<sup>th</sup> ed.). Los Angeles, CA: Author.
- Muthén, B. & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In R.D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 87-99). New York: Academic Press.
- Muthén, B. & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (Ed.), *Sociological methodology 1995* (pp. 216-316). Washington, DC: American Sociological Association.
- Muthén, B. & Satorra, A. (1996). Technical aspects of Muthén's LISOCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, 60, 489-503.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44, 443-460.
- Olsson, U., Drasgow, F. e Dorans, N.J. (1982). The polyserial correlation coefficient. *Psychometrika*, 47, 337-347.

- Olsson, U. H., Foss, T., Troye, S. V. & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modelling under conditions of misspecification and nonnormality. *Structural equation modelling*, 7(4), 557- 595.
- Poon, W.Y. & Lee, S.-Y. (1992). Maximum likelihood and generalized least squares analysis of two-level structural equation models. *Statistics and Probability Letters*, 14, 25-30.
- Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, 54, 131-151.
- Searle, S.R., Casella, G. & McCulloch, C.E. (1992). *Variance components*. New York, Wiley.
- Rabe-Hesketh, S., Skrondal, A. e Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69, 167-190.
- Rabe-Hesketh, S., Skrondal, A. e Zheng, X. (2007). Multilevel structural equation modelling. In S.Y. Lee (Ed.), *Handbook of Latent Variable and Related Models* (pp. 209-227). Amsterdam, Holland: Elsevier.
- Raudenbush, S.W. & Bryk, A. (1995). Maximum likelihood estimation for unbalanced multilevel covariance structure models via the EM algorithn. *British Journal of Mathematical and Statistical Psychology*, 48, 359-370.
- Raudenbush, S.W. & Bryk, A. (2002). *Hierarchical Linear Models. Second edition*. Thousand Oaks. Sage Publications, Inc.
- Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. In P. Marsden (Ed.), *Sociological methodology 1992* (pp. 249-278). Oxford, England: Blackwell Publishers.
- Satorra, A. & Bentler, P.M. (1994). Corrections to test statistics and standard errors on covariance structure analysis. *Psychometrika*, 66, 507-512.
- Searle, S.R., Casella, G. & McCulloch, C.E. (1992). *Variance components*. New York: Wiley.
- Skrondal, A. e Rabe-Hesketh, S. (2004). *Generalized latent variable modelling: Multilevel, Longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.
- Snijders, T.A. & Bosker, R. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

- Stapleton, L.M. (2006). Using multilevel structural equation modelling with complex sample data. In G. Hancock & R.O. Mueller (Ed.), *Structural equation modelling: A second course* (pp. 345-383). Greenwich, CT: Information Age Publishing.
- Yu, C. & Muthén, B. (2002). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Yuan, K.H. e Hayashi, K. (2005). On Muthén's maximum likelihood for two-level covariance structure models. *Psychometrika*, 70, 147-167.

# Appendice A

## La situazione occupazionale dei dottori di ricerca in Italia: una breve rassegna

### *Introduzione*

La rassegna sulle indagini riguardanti la condizione occupazionale dei dottori di ricerca riportata di seguito, peraltro non esaustiva, è stata fatta distinguendo innanzitutto le iniziative a livello internazionale da quelle nazionali. Queste ultime sono invece state distinte in base al soggetto proponente che, in alcuni casi, è diverso dal soggetto che ha realizzato l'indagine. Per ciascuna iniziativa, laddove possibile, sono state riportate le informazioni necessarie a caratterizzarle, cercando di assicurare una certa omogeneità e completezza rispetto a quanto è stato possibile raccogliere.

### **A.1 Panorama internazionale**

A livello internazionale tra i contributi che hanno avuto come oggetto gli esiti occupazionali dei dottori di ricerca si segnalano alcune iniziative. Martinelli (1999) esamina gli sbocchi occupazionali dei dottori di ricerca francesi. Nerad e Cerny (1999) analizzano gli esiti occupazionali dei dottori di ricerca negli Stati Uniti con riferimento all'indagine "Ph.D.'s – Ten Years Later". Mangematin (2000) valuta il mercato del

lavoro per i dottori di ricerca usando i dati di un'indagine sui dottori di ricerca di un ateneo francese. Enders (2002) esamina gli esiti occupazionali dei dottori di ricerca in Germania. Auriol (2007) confronta gli esiti occupazionali dei dottori di ricerca in sette paesi dell'OECD. Western *et al.* (2007) analizzano gli esiti occupazionali dei dottori di ricerca in Australia. Raddon e Sung (2009) sintetizzano la letteratura e gli studi empirici relativi ai dottori di ricerca britannici.

Da un punto di vista della tipologia di indagine, lo sviluppo della carriera si può studiare in modo ottimale ricorrendo ad un'indagine longitudinale su una o più coorti di dottori. Tuttavia, la realizzazione di un'indagine longitudinale è talmente complessa e costosa che usualmente l'evoluzione temporale viene analizzata confrontando gli esiti osservati in alcune coorti di dottori, come nell'indagine di Enders (2002) o come descritto nella rassegna di Raddon e Sung (2009).

L'articolo di Enders (2002) merita di essere descritto con qualche dettaglio perché riporta i risultati di una ricerca ampia e ben condotta. L'autore utilizza i dati di un'indagine sui dottori di ricerca tedeschi in sei discipline (biologia, economia, ingegneria elettrica, letteratura tedesca, matematica e scienze sociali). L'indagine, condotta nel 1999, si riferisce a tre coorti (1979-1980, 1984-1985 e 1989-1990), in modo da analizzare i cambiamenti e gli esiti nel lungo periodo. Il questionario, a cui hanno risposto 2244 dottori di ricerca (tasso di risposta del 52%), si compone di cinque aree tematiche: formazione durante il dottorato; transizione verso il lavoro; carriera e mobilità lavorativa; occupazione attuale; curriculum degli studi e caratteristiche socio-demografiche. Al fine di studiare le differenze in termini di occupazione tra dottorato di ricerca e laurea, sono stati intervistati anche 1895 laureati nelle stesse sei discipline. Venendo ai principali risultati, Enders (2002) rileva che durante gli studi molti dottori di ricerca svolgono un'attività lavorativa, che circa un terzo di loro mantiene anche successivamente. Il processo di acquisizione di un nuovo lavoro riguarda invece circa due terzi dei dottori e avviene in modo piuttosto rapido, poiché generalmente si conclude entro un anno dal conseguimento del titolo (con tempi appena più lunghi per letteratura tedesca e scienze sociali). In molti casi il primo lavoro non è quello definitivo, ma spesso determina, nel bene o nel male, gli sviluppi futuri. Ad esempio, il settore del primo impiego di solito rimane invariato nel lungo periodo. Inoltre, solo una minoranza dei dottori che accettano un primo lavoro modesto riescono successivamente

ad investire a tendenza e ottenere un lavoro pienamente soddisfacente. Dopo un anno dal conseguimento del titolo il settore di occupazione prevalente è quello pubblico per i dottori in biologia, letteratura tedesca, matematica e scienze sociali (circa il 60%), ma non per i dottori in economia e ingegneria elettrica (circa il 30%). A parità di disciplina, l'occupazione nel settore pubblico è più frequente per le femmine. Inoltre, comparando le tre coorti del campione, il ruolo del settore pubblico sembra diminuire negli anni, soprattutto per i dottori in letteratura e matematica (per i quali l'impiego nel settore pubblico è spesso un impiego da insegnante).

Lo sbocco naturale per un dottore di ricerca dovrebbe essere il settore dell'alta formazione o quello della Ricerca e Sviluppo: ciò accade per quasi due terzi dei dottori in ingegneria elettrica e matematica, ma per meno della metà dei dottori nelle altre discipline considerate (questa situazione è simile nelle tre coorti). Il livello di coerenza tra formazione ricevuta e mansioni svolte, valutato dai rispondenti su scala ordinale, è alto per l'83% di coloro che lavorano nell'alta formazione e per il 62% di coloro che lavorano nella Ricerca e Sviluppo. Al di fuori di questi settori la coerenza è alta per circa la metà dei dottori di ricerca (addirittura in due casi su tre l'intervistato afferma che una laurea sarebbe stata più appropriata). Questo non significa, però, che i dottori non lavorano nell'ambito dell'alta formazione o della ricerca siano poco soddisfatti del loro lavoro: infatti, coloro che si dichiarano insoddisfatti sono appena il 15%. Uno dei motivi di questo apparente paradosso va ricercato nel fatto che le posizioni apparentemente inappropriate sono spesso scelte volontariamente per motivi professionali (lavoro interessante, o con buone prospettive di carriera o con buona retribuzione) o personali (es. vicinanza alla famiglia). Per quanto riguarda il confronto tra dottori di ricerca e laureati, il dottorato sembra fornire maggiori opportunità in termini di probabilità di trovare lavoro e, nel lungo periodo, di fare carriera, ma le differenze in termini salariali sono modeste. Comunque, per molti aspetti i comportamenti e i destini dei dottori di ricerca nel mercato del lavoro sono simili a quelli dei laureati della stessa disciplina, per cui le maggiori differenze non sono tra dottori di ricerca e laureati, ma tra individui che hanno un certo tipo di formazione piuttosto che un altro (in particolare, la differenza più marcata in termini di esiti occupazionali si nota tra coloro che hanno una formazione ingegneristica e coloro che hanno una formazione letteraria).

## **A.2 Panorama nazionale**

Negli ultimi anni non sono mancate iniziative tese all'accertamento della condizione occupazionale dei dottori di ricerca ma il quadro che emerge non può essere considerato significativo. Infatti, si tratta nella generalità dei casi di iniziative che hanno interessato singoli atenei o specifici corsi di dottorato, e sono state realizzate soprattutto dai Nuclei di Valutazione Interna delle università. La loro validità è quindi limitata ai casi analizzati e non può essere considerata rappresentativa della situazione nazionale.

### **A.2.1 Iniziative MIUR e CNVSU**

#### *A.2.1.1 Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) e Ministero del Lavoro, della Salute e delle Politiche Sociali (MLSPS)*

Il 26 giugno 2003, il Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) e il Ministero del Lavoro, della Salute e delle Politiche Sociali (MLSPS) hanno siglato un accordo per promuovere un'indagine retrospettiva sui percorsi di studio e di lavoro di giovani che hanno partecipato a percorsi di alta formazione (dottorati di ricerca e master) cofinanziati dal Fondo Sociale Europeo (FSE) nelle Regioni Obiettivo 1<sup>110</sup> nell'ambito del Programma Operativo Nazionale (PON) "Ricerca, sviluppo tecnologico e alta formazione" programmazione 2000-2006. A seguito di tale intesa, si è costituito un gruppo di lavoro congiunto – Ministero dell'Istruzione, dell'Università e della Ricerca, Ministero del Lavoro della Salute e delle Politiche Sociali e Struttura nazionale di valutazione FSE ISFOL. L'indagine è stata realizzata in due fasi successive.

La prima fase dell'indagine è stata condotta a 6 mesi dalla conclusione degli interventi formativi. I risultati sono confluiti in un rapporto presentato al Comitato di Sorveglianza del PON "Ricerca, sviluppo tecnologico e alta formazione" dell'8 giugno 2007 che è stato successivamente pubblicato<sup>111</sup>. La rilevazione ha interessato un campione di 5.000 laureati che tra il 1999 e il 2004 hanno fatto richiesta di partecipare ad interventi formativi post-lauream finanziati dal FSE, indipendentemente dalla loro effettiva partecipazione alle attività stesse. Ciò al fine di poter valutare l'efficacia degli

---

<sup>110</sup> Le regioni interessate dall'intervento sono: Basilicata, Campania, Calabria, Puglia, Sardegna e Sicilia.

<sup>111</sup> ISFOL – Struttura nazionale di valutazione FSE. Gli esiti occupazionali dell'alta formazione nel Mezzogiorno, Collana I libri del Fondo Sociale Europeo, 2007.



interventi di alta formazione posti in essere, comparando gli esiti occupazionali dei “formati” con quelli dei soggetti che non hanno partecipato alla formazione, il cosiddetto “gruppo di controllo”. Nel complesso sono state eseguite 3.507 interviste, di cui 2.591 ai “formati” e 916 ai soggetti facenti parte del “gruppo di controllo”.

La seconda fase dell’indagine è stata realizzata allo scopo di consentire una ulteriore e più completa verifica degli esiti occupazionali conseguiti e la loro evoluzione nel tempo. L’indagine è stata quindi effettuata a 18 mesi dalla fine dei percorsi di alta formazione sugli stessi soggetti già contattati nella prima fase. I due Ministeri hanno quindi rinnovato l’incarico alla Struttura Nazionale di Valutazione FSE dell’ISFOL che ha proceduto ad un supplemento di indagine.

L’obiettivo dell’indagine era quello di valutare l’efficacia della formazione erogata a 18 mesi e di proporre un confronto con i dati rilevati a 6 mesi, evidenziando le maggiori differenze in termini occupazionali, in termini di tutela contrattuale e di reddito. Il rapporto offre un’analisi e una valutazione del loro inserimento lavorativo in termini non solo quantitativi ma anche qualitativi, approfondisce le caratteristiche del lavoro trovato, gli aspetti contrattuali e retributivi e le transizioni nella condizione occupazionale. Una riflessione viene inoltre sviluppata sul tema della formazione del capitale umano di eccellenza, analizzata in termini di qualità.

La popolazione di riferimento era composta da 3.241 individui, di cui 1.378 riguardano il dottorato e 1.863 i corsi post laurea. La rilevazione dei dati, effettuata nel corso del 2006, a 18 mesi di distanza dal conseguimento del titolo, è stata realizzata mediante interviste telefoniche. Le interviste sono state effettuate in parte con il metodo CATI (Computer assisted telephone interviewing) e in parte con il metodo CAWI (Computer assisted wireless interviewing).

Nel complesso sono state condotte in questa fase 2.000 interviste complete, di cui 814 (59,1%) per il dottorato e le restanti 1.186 (63,7%) per i corsi post laurea. Il tasso di risposta iniziale, che misura la proporzione di individui effettivamente intervistati rispetto a quelli appartenenti al campione, è stato relativamente basso (62,4%) rispetto alla prima indagine.

#### *A.2.1.2 Comitato Nazionale di Valutazione del Sistema Universitario (CNVSU)*

Il Comitato Nazionale di Valutazione del Sistema Universitario (CNVSU), alla fine del

2003 ha approvato e finanziato il “Progetto per la ricognizione, raccolta e analisi dei dati esistenti sul dottorato di ricerca e per l’indagine sull’inserimento professionale dei dottori di ricerca”. L’indagine è stata svolta da un gruppo di lavoro ADI (Associazione dottorandi e dottori di ricerca italiani)<sup>112</sup> con il supporto del dipartimento di Scienze Antropologiche di Genova. I risultati sono riportati nell’RdR 01/06<sup>113</sup> (rapporti di ricerca prodotti da altri per conto del Comitato).

Il progetto era composto da due fasi.

Nella prima fase il gruppo ha proposto un insieme di indicatori e variabili per il monitoraggio e la valutazione dei dottorati. Nella seconda fase è stata svolta l’indagine sull’inserimento professionale dei dottori di ricerca.

Gli obiettivi dell’indagine erano la valutazione della formazione ricevuta, la descrizione del profilo scientifico e professionale dei dottori di ricerca, la descrizione del percorso formativo e professionale seguito dopo il conseguimento del titolo, la caratterizzazione del segmento di mercato del lavoro svolto dai dottori di ricerca e la realizzazione e sperimentazione di metodologie e strumenti per la valutazione dei processi formativi di alta qualificazione (indicatori, variabili, strumenti e procedure di raccolta dati, modelli interpretativi).

La popolazione di riferimento era costituita dai dottori di ricerca che hanno conseguito il titolo nel periodo 1998-2003 nei 4 atenei che hanno aderito all’iniziativa: Pavia, Pisa, Siena e Salerno. I dottori di ricerca sono stati invitati dal loro ateneo a compilare un questionario on-line mediante identificativo e password personale. L’indagine è stata condotta nel periodo 20 giugno - 1 ottobre 2005.

I rispondenti sono stati 464, pari al 19,5% dei 2.383 dottori appartenenti alla popolazione di riferimento. Il tasso di risposta molto basso e la popolazione troppo limitata non consentono di far assumere all’indagine una validità generale.

---

<sup>112</sup> Il gruppo di lavoro era composto da Marco Bianchetti, Francesca della Ratta, Monica Lanzoni, Vincenzo Pischedda, Rocco Rizzo e Maria Carmen Usai.

<sup>113</sup> Il report sintetico è disponibile sul sito del CNVSU tra i documenti del comitato relativi all’anno 2006 (<http://www.cnvsu.it/publidoc/comitato/default.aspde>). Il report completo è invece disponibile sul sito dell’ADI <http://www.dottorato.it/qualita/>.

### **A.2.2 Iniziative di ateneo**

Le indagini sono state realizzate principalmente dai Nuclei di Valutazione dei diversi atenei, più raramente da altri servizi o strutture interne o esterne all'ateneo. I nuclei di valutazione svolgono ogni anno una relazione sullo stato dei corsi di dottorato attivi e una relazione sull'accREDITamento delle scuole e dei corsi di dottorato di ricerca per l'attivazione dei cicli ma, ad eccezione dei casi riportati di seguito, non svolgono indagini sugli esiti occupazionali dei dottori di ricerca.

#### *A.2.2.1 Nuclei di Valutazione delle Università di Milano, Milano Bicocca e Trento*

L'indagine è stata progettata congiuntamente dai nuclei di valutazione dei tre atenei e riguarda gli esiti occupazionali dei dottori di ricerca che hanno ottenuto il titolo tra il 1998 e il 2005, compresi i dottori dei cicli di studi VIII e XVII. La popolazione di riferimento era costituita da 1784 dottori di ricerca che hanno conseguito il titolo nelle tre Università nel periodo considerato. I rispondenti sono stati 1221 soggetti (68,4%). La rilevazione è stata effettuata dal Centro Cati del Dipartimento di Sociologia e Ricerca Sociale dell'Università di Milano Bicocca secondo la metodologia CATI (Computer Assisted Telephone Interviewing).

#### *A.2.2.2 Nucleo di valutazione della LUISS*

L'indagine è stata realizzata nel 2007 ed aveva la finalità analizzare il destino professionale, in termini di esiti occupazionali e di studio, dei dottori di ricerca ed affianca quella già condotta periodicamente dal Nucleo di Valutazione sull'inserimento professionale dei laureati. La popolazione di riferimento era costituita da coloro che, alla fine del 2006, avevano frequentato e terminato tutti i corsi di dottorato di cui la LUISS era sede amministrativa dal 1995, quando i corsi sono stati istituiti. La numerosità assoluta, era pari a 75 soggetti, 60 dei quali sono stati raggiunti telefonicamente (80,0%). La rilevazione è stata effettuata mediante intervista telefonica attraverso un questionario semi strutturato (domande "chiuse" e domande "aperte").

#### *A.2.2.3 Nucleo di valutazione dell'Università di Bologna*

L'indagine sugli esiti occupazionali dei dottori di ricerca dell'ateneo di Bologna è stata commissionata al Consorzio Interuniversitario ALMALAUREA su richiesta del

Magnifico Rettore e del Presidente del Nucleo di Valutazione dell'Ateneo di Bologna. L'intento dell'ateneo era quello di realizzare un sistema di rilevazione che consentisse di investigare, con continuità e in dettaglio, i destini occupazionali dei dottori di ricerca negli anni successivi al conseguimento del titolo. Tale obiettivo ha trovato riscontro nel progetto, già in fase avanzata, del Consorzio Interuniversitario AlmaLaurea di realizzare il monitoraggio per i dottori di ricerca di tutti gli Atenei consorziati.

L'indagine ha coinvolto 1.492 dottori, 565 dell'anno 2007, 527 del 2005 e 400 del 2003, intervistati, rispettivamente, ad uno, tre e cinque anni dalla conclusione del loro percorso di alta formazione.

L'indagine è stata realizzata attraverso interviste telefoniche secondo il metodo CATI (Computer Assisted Telephone Interviewing), svolte dal 6 maggio al 3 giugno 2008 dalla società SWG. L'interesse dei dottori nei confronti dell'indagine è stato elevato, se si considera che il tasso di risposta ottenuto è stato superiore, complessivamente, all'80,0% (84% ad un anno, oltre 78% a tre e cinque anni). Il questionario utilizzato per le interviste, sviluppato a partire dall'esperienza decennale acquisita nel campo delle indagini sull'inserimento professionale dei laureati dal Consorzio AlmaLaurea, e la professionalità degli operatori della società di rilevazione SWG, hanno favorito il buon esito della fase di rilevazione.

#### *A.2.2.4 Nucleo di valutazione dell'Università di Pavia*

Il Nucleo di valutazione dell'Università di Pavia ha realizzato due indagini sulla soddisfazione e gli sbocchi professionali dei dottori di ricerca dell'Ateneo di Pavia, la prima nel 2007, la seconda nel 2008. L'obiettivo delle indagini era quello di tracciare un primo bilancio dell'esperienza formativa e di raccogliere informazioni sui percorsi lavorativi che hanno caratterizzato il periodo successivo al conseguimento del titolo. Per raggiungere questi obiettivi, sono state raccolte le opinioni e le percezioni dei dottori di ricerca intervistati sulle tematiche della soddisfazione rispetto al corso di dottorato (attività didattica frequentata, attività di ricerca, attività didattica svolta, strutture e attrezzature), sui tempi e le modalità di inserimento occupazionale e sulla soddisfazione rispetto al lavoro svolto.

L'indagine del 2007 aveva una popolazione di riferimento costituita dai dottori di ricerca che avevano conseguito il titolo negli anni 2004, 2005 e 2006, pari

rispettivamente a 79, 91 e 90 unità. Il tasso di risposta complessivo è stato del 55,0%.

L'indagine del 2008, condotta via web nel periodo compreso tra maggio a settembre 2008, aveva una popolazione di riferimento costituita dai dottori di ricerca che avevano conseguito il titolo nel 2007, pari a 122 unità. In questo caso il tasso di risposta ha raggiunto il 71,8%.

#### *A.2.2.5 Nucleo di valutazione dell'Università di Camerino*

L'Università degli Studi di Camerino, a partire dall'anno 2005, ha avviato un'indagine per valutare l'impatto che il dottorato di ricerca ha nei confronti del mondo del lavoro e sugli sbocchi occupazionali dei neo dottori di ricerca. L'indagine viene realizzata dal Nucleo di Valutazione attraverso interviste telefoniche. Il questionario è articolato in tre sezioni (dati strutturali del dottore, sbocchi occupazionali e valutazione complessiva dell'esperienza da dottorando).

L'indagine del 2005 è stata condotta sulla popolazione dei “dottori di ricerca” dall'anno 2000, costituita da 65 unità distribuite in 5 cicli e 11 corsi di dottorato di ricerca. Le interviste telefoniche valide sono state 24, pari al 36,9% della popolazione di riferimento.

L'indagine del 2006 è stata condotta sulla popolazione dei “dottori di ricerca” ad uno e due anni dal conseguimento del titolo ed era costituita dai dottori licenziati nel 2003 e 2004: 45 unità, distribuite in 3 cicli e 10 corsi di dottorato di ricerca. Le interviste telefoniche valide sono state 31 pari, al 68,9% della popolazione di riferimento.

L'indagine del 2007 è stata condotta sulla popolazione dei “dottori di ricerca” a un anno dal conseguimento del titolo, quindi i dottori licenziati nel 2003 e 2004 sono stati intervistati a gennaio 2005 e 2006 mentre i dottori licenziati nel 2005 sono stati intervistati a gennaio 2007: 86 unità, distribuite in 4 cicli e 12 corsi di dottorato di ricerca. Le interviste telefoniche valide sono state 63, pari al 72,3% della popolazione di riferimento.

L'indagine del 2008 è stata condotta sulla popolazione dei “dottori di ricerca” ad un anno dal conseguimento del titolo, quindi i dottori licenziati nel 2004 e 2005 e 2006 sono stati intervistati rispettivamente a gennaio 2006, 2007 e 2008: 90 unità, distribuite in 3 cicli e 10 corsi di dottorato di ricerca. Le interviste telefoniche valide nei 3 anni

sono state 68, pari al 75% della popolazione di riferimento.

#### *A.2.2.6 Comitato Statistico dell'ateneo di Ferrara*

Il Comitato Statistico dell'ateneo di Ferrara ha realizzato due indagini di tipo campionario sugli sbocchi occupazionali dei dottori di ricerca del proprio ateneo, la prima realizzata nel 2005, la seconda nel 2007. In particolare, gli aspetti indagati sono stati la condizione occupazionale, la relazione tra lavoro e formazione universitaria e la soddisfazione per il lavoro svolto.

L'indagine realizzata nel 2005 aveva una popolazione di riferimento costituita da 288 dottori di ricerca con titolo conseguito dal 2001 al 21 aprile 2004. Il campione era costituito da 120 dottori di ricerca, 30 estratti casualmente da ciascuna delle 4 coorti annuali (2001, 2002, 2003, 2004) con stratificazione per area dottorato (allocazione proporzionale). La rilevazione dei dati è stata fatta secondo il metodo CATI (Computer Assisted Telephone Interviewing).

L'indagine realizzata nel 2005 aveva una popolazione di riferimento costituita da 249 dottori di ricerca con titolo conseguito nel 2004 e nel 2006. Gli obiettivi dell'indagine erano sostanzialmente gli stessi della precedente anche se nella seconda si è dato particolare risalto al confronto con i risultati ottenuti nell'indagine precedente. Il campione era costituito da 60 dottori di ricerca estratti casualmente dalle due coorti (2004 e 2006), con stratificazione per macroarea di dottorato (allocazione proporzionale). In particolare, la coorte 2004 è stata rilevata a 3 anni dal conseguimento del titolo mentre la coorte 2006 ad un anno. La rilevazione dei dati è stata fatta nell'aprile 2007 secondo il metodo CATI.

#### *A.2.2.7 Nucleo di Valutazione dell'Università Cattolica del Sacro Cuore di Milano*

Il Nucleo di Valutazione di Ateneo, ha promosso e condotto due indagini sugli sbocchi occupazionali di coloro che hanno conseguito il titolo di dottore di ricerca presso l'Università Cattolica. La prima, realizzata nel 2005, ha riguardato i dottori di ricerca che hanno conseguito il titolo al termine dei cicli dal primo al quindicesimo (I-XV). La seconda, realizzata nel 2007, ha riguardato i dottori che hanno conseguito il titolo al termine dei cicli dal sedicesimo al diciottesimo (XVI-XVIII). L'obiettivo primario delle indagini era la conoscenza degli sbocchi occupazionali ed il giudizio sul livello di

formazione ricevuto.

L'indagine realizzata nel 2005 aveva una popolazione di riferimento costituita da 812 dottori di ricerca, valore di circa il 10% inferiore rispetto al numero accertato di nominativi. L'intervallo temporale di interesse dell'indagine partiva dal 1987, anno in cui hanno conseguito il titolo i frequentanti il primo ciclo di dottorato, al primo semestre 2004, durante il quale si sono addottorati gli ultimi frequentanti il quindicesimo ciclo. I rispondenti sono stati 365, pari al 45% del totale, anche se non tutti hanno risposto alla totalità delle domande proposte. La rilevazione è stata realizzata tra il mese di novembre del 2003 e il primo semestre del 2004. Data l'ampiezza della rilevazione, è stato fatto ricorso al metodo dell'autocompilazione del questionario. L'invio del questionario è stato accompagnato da una lettera a firma della Direzione di supporto al Nucleo di Valutazione, che comunicava le ragioni dell'iniziativa. L'organizzazione della rilevazione è stata curata direttamente dalla Direzione di Supporto al Nucleo di Valutazione.

L'indagine realizzata nel 2007 ha riguardato un campione di 370 dottori di ricerca, i rispondenti sono stati 285, con una risposta pari al 79,2%.

#### *A.2.2.8 Servizio Formazione alla Ricerca e Servizio Ricerca dell'Università di Padova*

L'indagine sugli esiti professionali dei dottori di ricerca realizzata dal Servizio Formazione alla Ricerca e Servizio Ricerca dell'Università di Padova<sup>114</sup> si inserisce in un quadro di valutazione delle iniziative di Ateneo per la formazione alla ricerca. Tale valutazione è stata avviata nel 2001 con la prima indagine sugli esiti professionali degli assegnisti di ricerca (prendendo in considerazione tutti i contratti stipulati a partire dal 1999) ed è continuata nel 2003 con le seguenti indagini volte a esaminare gli esiti professionali dei dottori di ricerca (che avevano iniziato la loro formazione dottorale nei cicli XIII-XV), gli esiti professionali degli assegnisti di ricerca (che avevano concluso i contratti tra la fine del 2001 e la fine del 2003) e gli esiti professionali dei borsisti post-dottorato (che avevano ottenuto le borse nei bandi 1998-2001). Più in particolare, l'indagine si inserisce alla conclusione del primo triennio di sperimentazione delle Scuole di Dottorato avviate nel 2005 e rappresenta un quadro di sintesi che consente una prima valutazione delle azioni intraprese nella formazione alla ricerca dall'Ateneo

---

<sup>114</sup> L'Ufficio Dottorati di ricerca fa capo al servizio Formazione alla Ricerca dell'Ateneo.

di Padova. La rilevazione effettuata nel 2008 è stata promossa dai Delegati del Rettore alla Formazione alla Ricerca su invito del Magnifico Rettore.

L'iniziativa aveva una pluralità di obiettivi, la valutazione degli esiti professionali conseguiti dai dottori di ricerca, la stima del numero di dottori di ricerca che operano all'estero, la valutazione dell'utilizzo delle competenze acquisite durante il dottorato di ricerca e la valutazione ex post della soddisfazione dei dottori di ricerca.

La popolazione, composta da 1496 dottori di ricerca, è stata preliminarmente informata dell'avvio e degli scopi dell'indagine mediante una lettera di presentazione del Delegato del Rettore e del Responsabile del Servizio Ricerca. I dottori di ricerca coinvolti erano coloro che avevano terminato il dottorato negli anni 2004, 2005, 2006 e 2007 (cicli XVI – XIX, gli ultimi prima del passaggio da Corsi a Scuole di Dottorato). In questi anni erano attivi 86 corsi di dottorato.

I rispondenti all'indagine sono stati 1078, pari al 72,0% della popolazione, rappresentando in maniera significativa ed omogenea tutte le 7 macroaree (Matematica e fisica, Chimica e scienze della terra, Biologia, agraria e medicina veterinaria, Medicina e psicologia, Ingegneria, Discipline umanistiche e pedagogia e Discipline sociali).

La rilevazione è stata effettuata tramite un questionario elettronico con il metodo CAWI (Computer Assisted Web Interviewing) integrata con metodologie di tipo CATI (Computer Assisted Telephone Interviewing), con una struttura analoga alla precedente indagine del 2003. Le interviste sono state svolte nel mese di maggio 2008.

#### *A.2.2.9 Università di Siena*

L'indagine si inserisce nell'ambito del Progetto “Valutazione e meta-valutazione del sistema universitario: metodi, indicatori e modelli statistici” finanziato dal piano di Ateneo per la Ricerca dell'Università di Siena ed è stata curata dal prof. Giulio Ghellini. La finalità dell'indagine era la ricostruzione delle principali esperienze di vita (in particolare di lavoro e di alta formazione) che hanno caratterizzato il periodo successivo al completamento della esperienze formative del dottorato di ricerca, con particolare attenzione al ruolo giocato dalla formazione ricevuta. L'indagine era di tipo retrospettivo e censuario.

La popolazione di riferimento complessiva era costituita dai dottori di ricerca dei



cicli dal XV al XVIII, (titolo conseguito entro gennaio 2006) ed aveva una numerosità di 760 unità. La popolazione di riferimento considerata ai fini dell'indagine era, invece, di 726 unità perché sono stati esclusi i 34 dottori di nazionalità non italiana.

La rilevazione dei dati è stata effettuata nei mesi di giugno e luglio del 2007 secondo il metodo CAWI (Computer Assisted Web Interviewing) ed è stata organizzata prevedendo innanzitutto un primo invio e due successivi solleciti, tutti da effettuare via posta elettronica, in cui si invitavano i dottori di ricerca alla compilazione del questionario CAWI (Computer Assisted Web Interviewing). Oltre a questi, era programmato un ulteriore tentativo di contatto telefonico per i non rispondenti ed infine l'invio tramite posta di una lettera di invito alla compilazione per i soggetti che ancora non avevano risposto.

Il numero totale dei rispondenti è stato di 565, il tasso di risposta complessivo è stato di poco inferiore all'80% ed è risultato decisamente più elevato per i dottori dei cicli più recenti (si passa dal 70% del XV ciclo al 87% del XVIII).

### **A.2.3 Iniziative di altri enti**

#### *A.2.3.1 Indagine sui dottorati di ricerca in Campania*

L'indagine (cfr. Gallo, 2009) aveva la finalità di individuare i punti di forza e di debolezza della formazione dottorale erogata in Campania da parte dei cinque Atenei aderenti al progetto Uni Campania (all'ATS Uni Campania): l'Università degli Studi di Napoli - L'Orientale (U.N.O.), l'Università degli Studi di Napoli - Suor Orsola Benincasa (S.O.B.), l'Università degli Studi del Sannio (SANNIO), l'Università degli Studi di Napoli "Parthenope" (PARTHENOPE) e la Seconda Università degli Studi di Napoli (S.U.N.). In particolare, l'indagine era organizzata in due parti distinte, la prima per valutare la qualità dei servizi erogati dai dottorati di ricerca, la seconda il placement degli stessi. Relativamente al placement, l'obiettivo era quello di verificare se i percorsi formativi dei dottorati erano in grado di formare giovani ricercatori che potevano essere efficientemente integrati nel tessuto economico. Da questo punto di vista è stata posta particolare attenzione alla coerenza dell'occupazione rispetto al tempo impiegato per trovarla e la permanenza o meno nella Regione Campania. Il questionario prendeva in

esame alcuni aspetti in particolare, gli sbocchi professionali, la coerenza e le attività intraprese per la ricerca di un lavoro coerente con la formazione ricevuta.

La popolazione di riferimento era composta di 690 dottori, in quanto non è stato possibile ricevere i dati della S.U.N. relativamente ai cicli XV, XVI, XVII, XVIII e XIX in uno degli Atenei che aderisce al progetto Uni Campania. L'indagine, di tipo censuario.

Nel complesso sono stati raggiunti il 49,4% dei dottori di ricerca, di questi però 42 si sono rifiutati di rispondere, quindi i questionari raccolti sono stati 299, pari al 43,3% della popolazione.

La raccolta dei dati, iniziata il 15 gennaio 2008, ha riguardato gli ultimi cinque cicli dei dottorati svolti presso gli atenei che fanno parte di UniCampania. La strategia utilizzata per la rilevazione dei dati prevedeva la scelta tra la compilazione telefonica o via e-mail del questionario.

#### *A.2.3.2 Iniziativa Interuniversitaria STELLA (Statistiche sul TEma Laureati & Lavoro in Archivio on-line)*

L'Iniziativa Interuniversitaria STELLA svolge ormai da alcuni anni un'attività di monitoraggio dei percorsi formativi e lavorativi dei propri laureati ma ha deciso di affrontare per la prima volta un'indagine sui dottori di ricerca solo nel 2009. Tale indagine, considerata dagli autori stessi ancora sperimentale, aveva una duplice finalità, da un lato quella di completare il quadro conoscitivo delle indagini occupazionali dei laureati triennali, specialistici/magistrali e a ciclo unico, dall'altro quella di valutare i corsi di dottorato nell'ottica di supporto alla definizione delle relative strategie universitarie. I risultati sono stati presentati il giorno 7 luglio 2009. L'indagine è stata ripetuta nel 2010 ed i risultati sono stati presentati il 10 dicembre.

La prima indagine ha riguardato i dottori di ricerca di 7 dei 14 Atenei che già lo scorso anno avevano aderito all'iniziativa STELLA partecipando all'indagine sugli sbocchi occupazionali dei laureati: 4 del Nord (Università degli Studi di Bergamo, di Brescia, di Milano, di Milano-Bicocca) 2 del Centro (Università di Pisa e Scuola Superiore di Studi Universitari e Perfezionamento S. Anna di Pisa) e 1 del Sud (Palermo). L'Università di Pavia, aderente a STELLA, non ha partecipato alla rilevazione perché aveva già effettuato negli anni precedenti una rilevazione autonoma

sui propri dottori di ricerca.

La popolazione di riferimento era costituita da 3980 dottori di ricerca di nazionalità italiana che hanno conseguito il titolo negli anni solari 2005, 2006 e 2007.

La rilevazione dei dati, in forma censuaria, è stata realizzata in modalità CAWI (Computer Assisted Web Interviewing) nel primo trimestre 2009.

I rispondenti sono stati 1968, pari al 49,4% del totale. A causa del numero elevato di non rispondenti, peraltro non uniformemente distribuiti nei diversi atenei, sui non rispondenti CAWI (Computer Assisted Web Interviewing) è stato estratto un campione casuale stratificato (10%) per area disciplinare ed ateneo, con l'eccezione dell'Università del Studi di Milano Statale, ateneo che aveva ottenuto una partecipazione molto maggiore all'indagine web. A tali intervistati è stato chiesto se avessero ricevuto il questionario ed il motivo della mancata compilazione, oltre ad una descrizione dell'eventuale attività lavorativa e la possibilità di ripetere lo stesso percorso formativo.

La seconda indagine aveva caratteristiche analoghe alla precedente ma la popolazione di riferimento è costituita dai 2995 dottori di ricerca che hanno conseguito il titolo nel 2008 negli atenei facenti parte del consorzio. La rilevazione di tipo censuario mentre le interviste effettuate con modalità CAWI sono state 1579 (52,7%).

#### *A.2.3.3 ISTAT*

L'ISTAT ha svolto, nel corso del 2005, un'indagine pilota a livello nazionale sulle carriere dei dottori di ricerca. Tale esperienza può essere considerata la prima italiana finalizzata ad offrire una panoramica sull'inserimento professionale di due leve di dottori di ricerca (2000 e 2002) che hanno conseguito il titolo in una qualunque sede universitaria e settore disciplinare. In particolare, la finalità dell'indagine era quella di analizzare la loro condizione occupazionale e le esperienze di mobilità per lavoro a tre e a cinque anni di distanza dal conseguimento del titolo.

La popolazione di riferimento era costituita dai dottori di ricerca in tutte le sedi universitarie italiane che hanno conseguito il titolo negli anni solari 2000 e 2002, pari rispettivamente a 2903 e 4291 unità. L'indagine, di tipo campionario, prevedeva un disegno di campionamento a uno stadio stratificato per area disciplinare, ripartizione geografica, dimensione dell'ateneo e sesso (nel complesso 168 strati) per numerosità

campionaria pari a 1518, per i dottori del 2000 e 1535, per i dottori del 2002. Inoltre, per cogliere il maggior numero possibile di esperienze di mobilità internazionale, sono stati estratti due campioni indipendenti, uno per ciascuna leva di dottori stranieri censiti (circa il 2% del totale). L'indagine aveva come domini di stima il sesso, le 14 aree disciplinari di appartenenza dei corsi e le tre ripartizioni geografiche di collocazione degli atenei (Nord, Centro e Mezzogiorno). Il tasso di risposta complessivo è stato pari al 56,7%, leggermente più basso per i dottori del 2000 (55,4%) rispetto ai dottori del 2002 (58,4%). A circa tre anni dal conseguimento del titolo, il 92,8% dei dottori di ricerca del 2006 svolge un'attività lavorativa, il 5,4% è in cerca di occupazione, mentre l'1,8%, pur non lavorando, dichiara di non essere alla ricerca di lavoro.

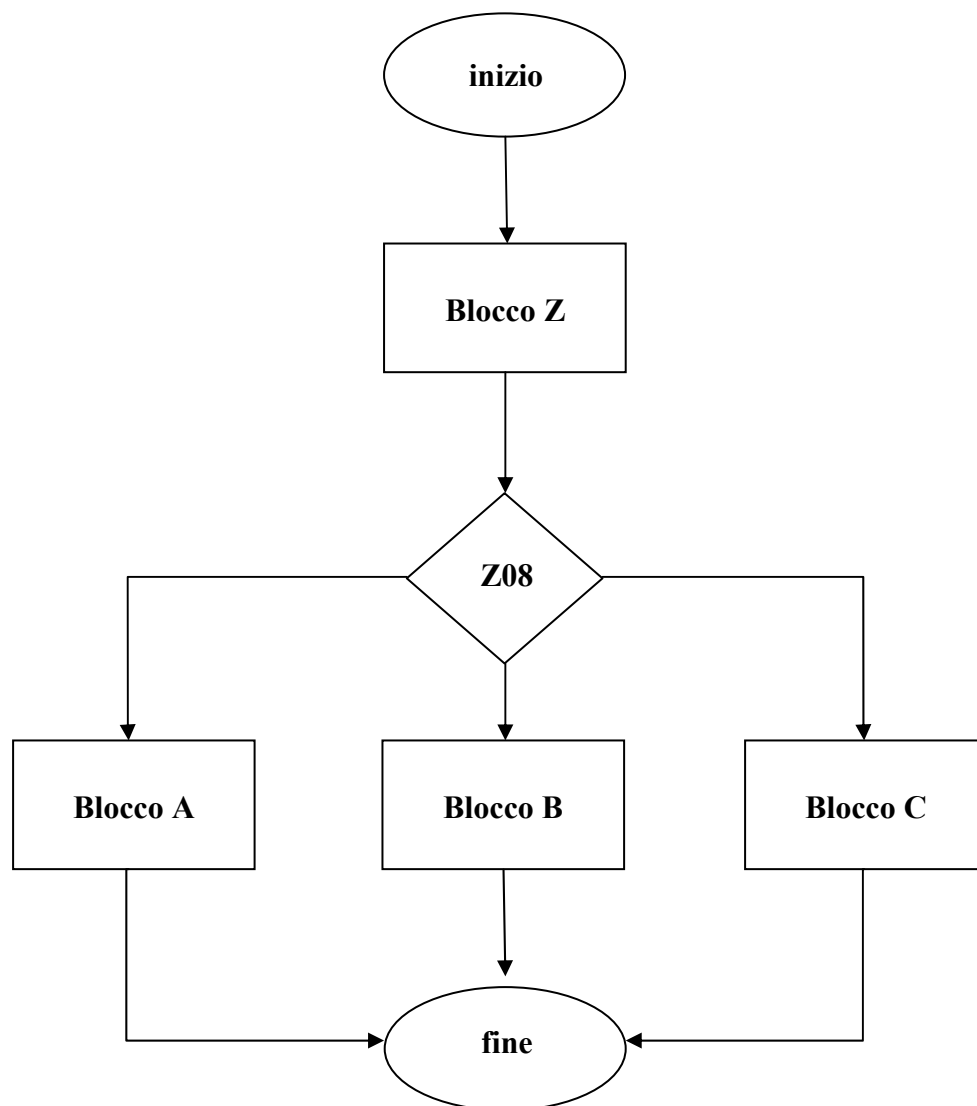
Tra dicembre 2009 e febbraio 2010 l'Istat ha svolto per la prima volta l'indagine sull'inserimento professionale dei dottori di ricerca, con l'obiettivo di conoscere i tempi e le modalità di collocamento nel mondo del lavoro di chi ha concluso un corso di dottorato di ricerca in un'università italiana. La popolazione di riferimento è costituita da circa 18600 dottori di ricerca che hanno conseguito il titolo negli anni solari 2004 e 2006, osservati rispettivamente a 3 e 5 anni dal conseguimento del titolo. Le interviste sono state realizzate secondo il metodo CATI (Computer Assisted Telephone Interviewing) ma è previsto un modulo in forma CAWI (Computer Assisted Web Interviewing) per gli aspetti più complicati.

I primi risultati sono stati presentati il 14 dicembre 2010. A circa tre anni dal conseguimento del titolo, il 92,8% dei dottori di ricerca del 2006 svolge un'attività lavorativa, il 5,4% è in cerca di occupazione, mentre l'1,8%, pur non lavorando, dichiara di non essere alla ricerca di lavoro. La quota di occupati tra i dottori di ricerca del 2004 (intervistati quindi a cinque anni dal conseguimento del dottorato) sale al 94,2% e quella di quanti sono ancora in cerca di lavoro scende al 4,4%, mentre appare simile quella relativa a coloro che non lavorano e non cercano lavoro (1,5%). Tra i dottori, sia a tre anni sia a cinque anni dal dottorato, si rileva una consistente quota di persone occupate da prima del conseguimento del titolo, rispettivamente il 29,7% e il 24,6%.

## **Appendice B**

Condizione occupazionale dei Dottori di  
Ricerca: questionario ad un anno dal  
conseguimento del titolo

## Struttura



\*\*\*\*\*

### **Domande del Blocco Z (tutti)**

\*\*\*\*\*

- Z00**      **Verifica dei dati d'archivio**  
Anno di nascita  
Genere  
Università di conseguimento del titolo di laurea  
Denominazione della laurea  
Anno di laurea e voto di laurea (compresa lode)  
Anno solare e mese di consegna della tesi  
Anno solare di conseguimento del titolo [fare riferimento all'esame finale] → ANNO  
Mese di conseguimento del titolo → MESE
- Z01**      **Qual è il motivo principale per cui ha deciso di iscriversi al Corso di Dottorato in cui ha conseguito il titolo di Dottore di Ricerca?**
- 1          per completare/arricchire la mia formazione
  - 2          per conseguire un titolo dal più alto potere professionalizzante
  - 3          interesse per lo studio
  - 5          per fare attività di ricerca
  - 6          per l'interesse nei confronti di una futura carriera accademica
  - 7          Altro
- Z02**      **Durante il Dottorato in cui ha conseguito il titolo di Dottore di Ricerca Lei è stato titolare di:**
- 1          un posto con borsa finanziata dall'Università / MIUR
  - 2          un posto con borsa finanziata da enti pubblici
  - 3          un posto con borsa finanziata da enti privati
  - 5          un posto senza borsa ma coperto da assegno di ricerca
  - 6          un posto senza borsa e non coperto da assegno di ricerca
  - 7          Altro
- Z03**      **Di seguito Le vengono elencati una serie di aspetti della formazione ricevuta durante il Corso di Dottorato. Per ciascuno mi dovrebbe dire quanto è soddisfatto, su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo).**
- |   |   |                       |
|---|---|-----------------------|
| 1 | contenuto degli insegnamenti                            | [1 ... 10] [no corsi] |
| 2 | qualità della didattica del personale docente           | [1 ... 10] [no corsi] |
| 3 | qualità scientifica di professori/ricercatori           | [1 ... 10]            |
| 4 | sviluppo / addestramento della capacità di fare ricerca | [1 ... 10]            |
| 5 | possibilità di pubblicare (articoli, volumi, ...)       | [1 ... 10]            |

**Z04      Se potesse tornare indietro si iscriverebbe di nuovo ad un Corso di Dottorato?**

- 1      sì, allo stesso Corso nella stessa sede universitaria
- 2      sì, allo stesso Corso ma in altra sede universitaria
- 3      sì ma ad altro Corso
- 4      No

**Z05      Durante il Corso di Dottorato ha compiuto periodi di studio all'estero?**

- 1      Sì
- 2      No

**Z06      Durante il Corso di Dottorato ha svolto una qualche attività lavorativa (inclusi co.co.co. e co.co.pro.)?**

- 1      Sì, attività attinente all'attività di ricerca del dottorato
- 2      Sì, attività non attinente all'attività di ricerca del dottorato
- 3      No

**Z07      Lei lavorava al momento del conseguimento del titolo di Dottore di Ricerca?**

- 1      Sì
- 2      No



**Z08** Può ricostruire il percorso compiuto dal conseguimento del titolo di Dottore di Ricerca ad oggi?

(faccia riferimento all'anno solare di conseguimento del titolo, ovvero all'anno in cui è stato sostenuto l'esame finale)

*Per ognuno dei momenti indichi la condizione prevalente.*

- 1 lavoro con assegno di ricerca, borsa post-doc o altra borsa
- 2 lavoro con altra forma contrattuale a tempo determinato (contratto formazione lavoro, collaborazione, altro contratto atipico, ecc.) oppure lavoro senza contratto
- 3 lavoro a tempo indeterminato / autonomo effettivo
- 4 attività di formazione post-dottorato retribuita
- 5 attività di formazione post-dottorato *non* retribuita
- 6 periodo di studio in preparazione a concorsi pubblici/esame di stato
- 7 ricerca di un lavoro
- 8 altra condizione di inattività (non lavora e non cerca lavoro) diversa dallo studio
- 9 altro
- 10 non ricorda (solo per domande Z08B e Z08C)

<b>Z08A</b>	<b>Z08B</b> (solo se ANNO = 2003 o ANNO = 1998)	<b>Z08C</b> (solo se ANNO = 1998)
Oggi	ad 1 anno dal conseguimento del titolo	a 5 anni dal conseguimento del titolo

**Z09** (solo se ((ANNO =1998) e Z08A >3 e Z08B >3 e Z08C >3) o (ANNO =2003 e Z08A >3 e Z08B >3) o (ANNO=2008 e Z08A>3))

Dopo il conseguimento del titolo di Dottore di Ricerca ha comunque svolto una qualche attività lavorativa?

- 1 Sì
- 2 No, non ho mai lavorato dopo il conseguimento del titolo

\*\*\*\*\*

**Domande del Blocco A: chi attualmente lavora**  
**(solo se Z08A <=3)**

\*\*\*\*\*

**A02** (solo se Z07 = 1)

**Il lavoro attuale è lo stesso che svolgeva al momento del conseguimento del titolo?** Non si considerino gli avanzamenti di carriera, né i trasferimenti (es. risponde “Si” chi era ricercatore di ruolo e diviene professore associato, risponde “No” chi era assegnista di ricerca e diviene ricercatore di ruolo)

- 1 Si (→ A05)  
2 No

**A03** (solo se Z08B <= 3 e (ANNO = 2003 o ANNO = 1998))

**Il lavoro attuale è lo stesso che svolgeva ad un anno dal conseguimento del titolo?** Non si considerino gli avanzamenti di carriera, né i trasferimenti (es. risponde “Si” chi era ricercatore di ruolo e diviene professore associato, risponde “No” chi era assegnista di ricerca e diviene ricercatore di ruolo)

- 1 Si (→ A05)  
2 No

**A04** (solo se Z08C <= 3 e ANNO = 1998)

**Il lavoro attuale è lo stesso che svolgeva a cinque anni dal conseguimento del titolo?** Non si considerino gli avanzamenti di carriera, né i trasferimenti (es. risponde “Si” chi era ricercatore di ruolo e diviene professore associato, risponde “No” chi era assegnista di ricerca e diviene ricercatore di ruolo)

- 1 Si  
2 No

**A05** (solo se Z07 = 1 e A02 = 1)

**Il conseguimento del titolo di Dottore di Ricerca ha comportato un miglioramento nel suo lavoro?**

- 1 Si  
2 No (→ A11)

**A06** (solo se Z07 = 1 e A02 = 1 e A05 = 1)

**Quali aspetti del suo lavoro sono migliorati?**

- |    |   |         |
|----|---|---------|
| 1  | Livello di retribuzione attuale                                 | (→ A11) |
| 2  | Prospettive di guadagno   | (→ A11) |
| 3  | Possibilità di carriera   | (→ A11) |
| 4  | Stabilità/sicurezza sul lavoro                                  | (→ A11) |
| 5  | Coerenza con studi fatti  | (→ A11) |
| 6  | Utilizzo delle competenze acquisite                             | (→ A11) |
| 7  | Acquisizione di professionalità                                 | (→ A11) |
| 8  | Rispondenza ai propri interessi culturali                       | (→ A11) |
| 9  | Indipendenza e autonomia sul lavoro                             | (→ A11) |
| 10 | Coinvolgimento nei processi decisionali                         | (→ A11) |
| 11 | Flessibilità dell'orario di lavoro                              | (→ A11) |
| 12 | Tempo libero  | (→ A11) |
| 13 | Localizzazione sede di lavoro                                   | (→ A11) |
| 14 | Rapporti coi colleghi   | (→ A11) |
| 15 | Miglioramento generale rispetto alle aspettative iniziali       | (→ A11) |
| 16 | Miglioramento generale rispetto alla attività lavorativa ideale | (→ A11) |

**A07** Dopo quanti mesi dal conseguimento del titolo ha trovato il primo lavoro iniziato dopo conseguimento del titolo di Dottore di Ricerca?

- 0    meno di un mese dopo il conseguimento del titolo  
...    numero di mesi

**A08** Il suo lavoro attuale coincide con il primo lavoro iniziato dopo conseguimento del titolo di Dottore di Ricerca?  
(non tenga conto degli eventuali avanzamenti di carriera)

- 1    Sì  
2    No

**A09** (solo se A08 = 2)

**Può indicare mese ed anno di inizio della sua attuale attività lavorativa?**

Mese (A09M), Anno (A09A)

**A10 In base a quale canale ha trovato l'attuale lavoro?**

- 1 lettura di offerte impiego sui giornali
- 2 risposta a inviti a partecipare a colloqui da parte di aziende
- 3 contatti con datori di lavoro su segnalazione università/docenti
- 4 contatti con datori di lavoro su iniziativa personale o altra iniziativa
- 5 domande o partecipazione a concorsi pubblici
- 6 proseguendo un'attività familiare già esistente
- 7 domande a provveditorati o presidi di insegnamento
- 8 tramite siti internet e/o motori di ricerca
- 9 Altro

**A11 Il suo attuale lavoro è di tipo autonomo o dipendente?**

- 1 autonomo (→ A15)
- 2 dipendente inclusi contratti di collaborazione (→ A12)

**A12 (solo se A11 = 2)**

**Con che tipo di contratto è regolato il lavoro da Lei attualmente svolto?**

- 1 tempo pieno e indeterminato
- 2 part-time, tempo indeterminato
- 3 tempo pieno, a termine
- 4 part-time, a termine
- 5 collaborazione coordinata e continuativa
- 6 collaborazione a progetto
- 7 collaborazione occasionale
- 8 prestazione d'opera
- 9 assegno di ricerca o borsa di studio
- 10 non è regolato da contratto
- 11 Altro

**A13 (solo se A11 = 2 e A12 ≤ 2)**

**Se ha fatto avanzamenti di carriera nell'ambito del lavoro attualmente svolto, da quanto tempo ricopre la posizione giuridica attuale? (ad es: in ambito universitario, se attualmente è professore associato indichi da quanto tempo ricopre tale ruolo)**

- ... numero di mesi (tra 0 e 132)
- 99 nessun avanzamento di carriera

**A14** (solo se A11 = 2)

**Può qualificarmi sotto il profilo giuridico il suo lavoro attuale?**

(coloro che svolgono la propria attività all'estero sono invitati a fare riferimento al profilo affine)

- |           |   |         |
|-----------|---|---------|
| <b>1</b>  | dirigente   | (→ A16) |
| <b>2</b>  | direttivo / quadro  | (→ A16) |
| <b>3</b>  | impiegato o intermedio                                    | (→ A16) |
| <b>4</b>  | ricercatore nel settore privato                           | (→ A20) |
| <b>5</b>  | assegnista / borsista                                     | (→ A18) |
| <b>6</b>  | ricercatore universitario a tempo determinato             | (→ A18) |
| <b>7</b>  | ricercatore universitario di ruolo                        | (→ A18) |
| <b>8</b>  | professore universitario associato                        | (→ A18) |
| <b>9</b>  | professore universitario ordinario                        | (→ A18) |
| <b>10</b> | ricercatore (settore pubblico non universitario)          | (→ A19) |
| <b>11</b> | primo ricercatore (settore pubblico non universitario)    | (→ A19) |
| <b>12</b> | dirigente di ricerca (settore pubblico non universitario) | (→ A19) |
| <b>13</b> | docente alle scuole medie inferiori                       | (→ A24) |
| <b>14</b> | docente alle scuole medie superiori                       | (→ A24) |

**A15** (solo se A11 = 1)

**Può qualificarmi sotto il profilo giuridico il suo lavoro attuale?**

- |          |   |
|----------|---|
| <b>1</b> | Imprenditore                            |
| <b>2</b> | Libero professionista iscritto albo     |
| <b>3</b> | Libero professionista non iscritto albo |

**A16** (solo se A14 ≤ 3 o A11 = 1)

**Nel suo lavoro svolge mansioni di ricerca e sviluppo?**

- |          |                                     |
|----------|-------------------------------------|
| <b>1</b> | Sì, come attività principale        |
| <b>2</b> | Sì, ma non come attività principale |
| <b>3</b> | No                                  |

**A17** (solo se A14 ≤ 3 o A11 = 1)

**Nel suo lavoro svolge attività di innovazione?**

- |          |   |         |
|----------|---|---------|
| <b>1</b> | Sì, sia innovazione di processo che innovazione di prodotto | (→ A24) |
| <b>2</b> | Sì, innovazione di processo                                 | (→ A24) |
| <b>3</b> | Sì, innovazione di prodotto                                 | (→ A24) |
| <b>4</b> | No  | (→ A24) |

- A18** (solo se  $A14 \geq 5$  e  $A14 \leq 9$ )
- Presso quale sede universitaria presta servizio?**
- 1** nello stesso Ateneo in cui ha svolto il dottorato
  - 2** In un Ateneo diverso ma nella stessa regione
  - 3** In un Ateneo diverso di altra regione italiana
  - 4** all'estero
- A19** (solo se  $(A14 \geq 4$  e  $A14 \leq 12)$  o  $A16 \leq 2$ )
- Le attività di ricerca svolte nella posizione attuale sono coerenti con la formazione acquisita nel Corso di Dottorato in cui ha conseguito il titolo?**
- 1** pienamente coerente
  - 2** parzialmente coerente
  - 3** non coerente
- A20A** (solo se  $(A14 \geq 4$  e  $A14 \leq 12)$  o  $A16 \leq 2$ )
- Indichi il numero di lavori da Lei pubblicati su riviste Scientifiche a diffusione nazionale dopo il conseguimento del titolo di Dottore di Ricerca**
- ... Numero
- A20B** (solo se  $(A14 \geq 4$  e  $A14 \leq 12)$  o  $A16 \leq 2$ )
- Indichi il numero delle Sue monografie pubblicate da editore nazionale:**
- ... Numero
- A21A** (solo se  $(A14 \geq 4$  e  $A14 \leq 12)$  o  $A16 \leq 2$ )
- Indichi il numero di lavori da Lei pubblicati su riviste Scientifiche a diffusione internazionale dopo il conseguimento del titolo di Dottore di Ricerca**
- ... Numero
- A21B** (solo se  $(A14 \geq 4$  e  $A14 \leq 12)$  o  $A16 \leq 2$ )
- Indichi il numero delle Sue monografie pubblicate da editore internazionale:**
- ... Numero
- A22** (solo se  $(A14 \geq 4$  e  $A14 \leq 12)$  o  $A16 \leq 2$ )
- A quanti progetti di ricerca nazionali ha partecipato dopo il conseguimento del titolo di Dottore di Ricerca**
- ... Numero

**A23** (solo se (A14 >= 4 e A14 <= 12) o A16 <= 2)

**A quanti progetti di ricerca internazionali (cioè finanziati da enti pubblici europei o comunque non italiani) ha partecipato dopo il conseguimento del titolo di Dottore di Ricerca**

... Numero

**A24** (solo se A14 < 4)

**Lavora in:**

- |          |   |         |
|----------|---|---------|
| <b>1</b> | Ente pubblico   | (→ A26) |
| <b>2</b> | Azienda a partecipazione pubblica                         |         |
| <b>3</b> | Azienda privata italiana (anche se in filiale all'estero) |         |
| <b>4</b> | Azienda privata straniera (anche se in filiale in Italia) |         |
| <b>5</b> | Organizzazione no-profit                                  |         |

**A25** (solo se A24 > 1)

**Quanti dipendenti ha l'azienda per cui lavora?**

- |          |                        |
|----------|------------------------|
| <b>1</b> | fino a 10 dipendenti   |
| <b>2</b> | da 11 a 50 dipendenti  |
| <b>3</b> | da 51 a 250 dipendenti |
| <b>4</b> | più di 250 dipendenti  |

**A26** (solo se A14 < 5)

**Può indicarmi il ramo di attività economica dell'azienda per cui lavora:**

- |           |                                     |           |  |
|-----------|-------------------------------------|-----------|--|
| <b>1</b>  | Agricol. - foreste - caccia - pesca | <b>14</b> | Commercio, Alberghi, Esercizi Pubbl.   |
| <b>2</b>  | Stampa ed editoria                  | <b>15</b> | Trasporti                              |
| <b>3</b>  | Energia, Gas, Acqua                 | <b>16</b> | Comunicazioni e Telecomunicazioni      |
| <b>4</b>  | Estrazione mineraria                | <b>17</b> | Credito e Assicurazioni                |
| <b>5</b>  | Chimica                             | <b>18</b> | Consulenza legale, amm.va, contabile   |
| <b>6</b>  | Metalmec. e Mecc. precisione        | <b>19</b> | Pubblicità, pubbliche relazioni        |
| <b>7</b>  | Alimentari e Tabacchi               | <b>20</b> | Informatica, elaborazione dati         |
| <b>8</b>  | Tessile, Abbigl, Cuoio, Calzature   | <b>21</b> | Servizi alle imprese                   |
| <b>9</b>  | Legno, Arredamento                  | <b>22</b> | Servizi ricreativi e culturali         |
| <b>10</b> | Carta, Gomma, Plastica              | <b>23</b> | Pubblica Amministrazione, Forze Armate |
| <b>11</b> | Costruz. - installa. d'impianti     | <b>24</b> | Istruzione, ricerca e sviluppo         |
| <b>12</b> | Elettronica e Elettrotecnica        | <b>25</b> | Sanità                                 |
| <b>13</b> | Manifattura varia                   | <b>26</b> | Altri Servizi Sociali                  |

**A27** Numero di ore settimanali dedicate abitualmente a questa attività lavorativa

... Numero di ore

**A28A** (solo se  $A11 = 1$  o  $A14 \leq 5$ )

**Il suo lavoro si svolge prevalentemente:**

- 1** in Italia (indicare regione e provincia)
- 2** all'estero (indicare stato)
- 3** in parte in Italia e in parte all'estero

**A28B** (solo se ( $A11 = 1$  o  $A14 \leq 5$ ) e  $A28A > 1$ )

**Qual è il motivo prevalente che l'ha indotta a lavorare all'estero?**

- 1** Mancanza di opportunità professionali in Italia rispetto al titolo conseguito
- 2** Mancanza di opportunità professionali in Italia rispetto alle proprie aspirazioni
- 3** Miglior livello di retribuzione
- 4** Motivi personali

**A29** **Quanto le è utile sul lavoro la formazione ricevuta durante il corso di dottorato?**

- 1** È fondamentale per le mansioni che svolgo
- 2** È utile per il mio approccio al lavoro, anche se non ne faccio un uso specifico
- 3** Ne faccio un uso piuttosto limitato
- 4** Non ne faccio alcun uso



**A30** **Può esprimere su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo) il suo livello di soddisfazione per i seguenti aspetti del lavoro attuale?**

**(Rispondere a tutti i quesiti)**

	Per niente		Moltissimo
Livello di retribuzione attuale	<b>1</b>	...	<b>10</b>
Prospettive di guadagno	<b>1</b>	...	<b>10</b>
Possibilità di carriera	<b>1</b>	...	<b>10</b>
Stabilità/sicurezza sul lavoro	<b>1</b>	...	<b>10</b>
Coerenza con studi fatti	<b>1</b>	...	<b>10</b>
Utilizzo delle competenze acquisite	<b>1</b>	...	<b>10</b>
Acquisizione di professionalità	<b>1</b>	...	<b>10</b>
Rispondenza ai propri interessi culturali	<b>1</b>	...	<b>10</b>
Indipendenza e autonomia sul lavoro	<b>1</b>	...	<b>10</b>
Coinvolgimento nei processi decisionali	<b>1</b>	...	<b>10</b>
Flessibilità dell'orario di lavoro	<b>1</b>	...	<b>10</b>
Tempo libero	<b>1</b>	...	<b>10</b>
Localizzazione sede di lavoro	<b>1</b>	...	<b>10</b>
Rapporti coi colleghi	<b>1</b>	...	<b>10</b>
Rispetto alle aspettative iniziali	<b>1</b>	...	<b>10</b>
Rispetto alle mansioni svolte	<b>1</b>	...	<b>10</b>
Rispetto al titolo di Dottore di Ricerca	<b>1</b>	...	<b>10</b>
E nel suo complesso	<b>1</b>	...	<b>10</b>

**A31** **Oltre all'attività principale Lei svolge un'altra attività retribuita?**

- 1**    Si  
**2**    No

**A32**      **Può indicare il reddito mensile netto medio da Lei percepito?**  
**Consideri tutte le attività retribuite.**

**(Le ricordiamo che i dati sono protetti dalla normativa sulla privacy e non possono essere divulgati, né comunicati a terzi)**

- 1**      meno di € 500
- 2**      € 500 - € 1000
- 3**      € 1001 - € 1500
- 4**      € 1501 - € 2000
- 5**      € 2001 - € 2500
- 6**      € 2501 - € 3000
- 7**      € 3001 - € 3500
- 8**      € 3501 - € 4000
- 9**      € 4001 - € 6000
- 10**    oltre € 6000

**A33**      **Negli ultimi tre mesi Lei ha cercato attivamente un nuovo lavoro?**

- 1**      Sì **(→ A34)**
- 2**      No **FINE**

**A34**      **(solo se A33 = 1)**

**Qual è la principale motivazione che l'ha indotta ad avviare la ricerca di un nuovo lavoro?**

- 1**      Stabilità del posto di lavoro
- 2**      Insoddisfazione per il livello di retribuzione
- 3**      Più elevata possibilità di carriera
- 4**      Maggiore coerenza con gli studi fatti
- 5**      Maggiore rispondenza al titolo posseduto
- 6**      Possibilità di acquisire maggiore professionalità
- 7**      Maggiore rispondenza ai propri interessi culturali
- 8**      Maggiore indipendenza e autonomia sul lavoro
- 9**      Localizzazione della sede di lavoro
- 10**    Per avere più tempo libero

**FINE**

\*\*\*\*\*

\*\*\*\*\*

### Domande del Blocco B:

*chi attualmente non lavora ma ha lavorato dopo il conseguimento del titolo*  
(solo se Z08A >3 e ((Z08B <=3 o Z08C <=3) o Z09 = 1))

\*\*\*\*\*

\*\*\*\*\*

**B01** (solo se Z07 = 1)

**L'ultimo lavoro svolto era lo stesso che svolgeva al momento del conseguimento del titolo?** Non si considerino gli avanzamenti di carriera, né i trasferimenti (es. risponde "Sì" chi era ricercatore di ruolo e diviene professore associato, risponde "No" chi era assegnista di ricerca e diviene ricercatore di ruolo)

- 1      Sì (→ B04)  
2      No

**B02** (solo se Z08B <= 3 e (ANNO = 2003 o ANNO = 1998))

**L'ultimo lavoro svolto era lo stesso che svolgeva ad un anno dal conseguimento del titolo?** Non si considerino gli avanzamenti di carriera, né i trasferimenti (es. risponde "Sì" chi era ricercatore di ruolo e diviene professore associato, risponde "No" chi era assegnista di ricerca e diviene ricercatore di ruolo)

- 1      Sì (→ B04)  
2      No

**B03** (solo se Z08C <= 3 e ANNO = 1998)

**L'ultimo lavoro svolto era lo stesso che svolgeva a cinque anni dal conseguimento del titolo?** Non si considerino gli avanzamenti di carriera, né i trasferimenti (es. risponde "Sì" chi era ricercatore di ruolo e diviene professore associato, risponde "No" chi era assegnista di ricerca e diviene ricercatore di ruolo)

- 1      Sì  
2      No

**B04** (solo se B01 = 1)

**Il conseguimento del titolo di Dottore di Ricerca aveva comportato un miglioramento nel suo lavoro?**

- 1      sì  
2      no

**B04b** (solo se B01 = 1 & B04 = 1)

**Quali aspetti del suo lavoro erano migliorati?**

- |    |  |         |
|----|--|---------|
| 1  | Livello di retribuzione attuale                        | (→ A11) |
| 2  | Prospettive di guadagno                                | (→ A11) |
| 3  | Possibilità di carriera                                | (→ A11) |
| 4  | Stabilità/sicurezza sul lavoro                         | (→ A11) |
| 5  | Coerenza con studi fatti                               | (→ A11) |
| 6  | Utilizzo delle competenze acquisite                    | (→ A11) |
| 7  | Acquisizione di professionalità                        | (→ A11) |
| 8  | Rispondenza ai propri interessi culturali              | (→ A11) |
| 9  | Indipendenza e autonomia sul lavoro                    | (→ A11) |
| 10 | Coinvolgimento nei processi decisionali                | (→ A11) |
| 11 | Flessibilità dell'orario di lavoro                     | (→ A11) |
| 12 | Tempo libero   | (→ A11) |
| 13 | Localizzazione sede di lavoro                          | (→ A11) |
| 14 | Rapporti coi colleghi                                  | (→ A11) |
| 15 | Miglioramento rispetto alle aspettative iniziali       | (→ A11) |
| 16 | Miglioramento rispetto alla attività lavorativa ideale | (→ A11) |

**B05** (solo se Z09 = 1)

**Dopo quanti mesi dal conseguimento del titolo di Dottore di Ricerca aveva trovato il primo lavoro svolto?**

- 0      meno di un mese dopo il conseguimento del titolo  
...    numero di mesi

**B06** **Per quale motivo è stata interrotta la sua ultima attività lavorativa?**

- |    |   |
|----|---|
| 1  | per iniziare un'attività di formazione post-dottorato   |
| 2  | Dimesso perché non soddisfatto del trattamento economico  |
| 3  | Dimesso perché non soddisfatto di quel tipo di lavoro/contratto   |
| 4  | Dimesso perché non soddisfatto della sede di lavoro   |
| 5  | Dimesso per motivi personali  |
| 6  | Licenziamento da parte del datore di lavoro o altra interruzione da lei <u>non voluta</u> (es. chiusura attività) |
| 7  | Scadenza regolare di un contratto a tempo determinato o termine di un lavoro occasionale o stagionale             |
| 8  | Maternità   |
| 9  | Per iniziare il servizio civile nazionale volontario  |
| 10 | Altro   |

**B07** Negli ultimi tre mesi Lei ha cercato attivamente un lavoro?

- 1 sì  
2 no

(→ B08)

**FINE**

**B08** **Z36** (solo se B07 = 1)

**\_SSSUP**

**Quale tipo di lavoro cerca?**

- |                                  |          |  |
|----------------------------------|----------|--|
| <b>LAVORATORE<br/>DIPENDENTE</b> | <b>1</b> | Dirigente  |
|                                  | <b>2</b> | Direttivo/Quadro                                   |
|                                  | <b>3</b> | Impiegato o intermedio                             |
|                                  | <b>4</b> | Ricercatore o Docente universitario                |
|                                  | <b>5</b> | Ricercatore nel settore pubblico non universitario |
|                                  | <b>6</b> | Docente alle Scuole Medie (Inferiori o Superiori)  |
| <b>AUTO<br/>NOMO</b>             | <b>7</b> | Lavoro autonomo                                    |
|                                  | <b>8</b> | Libero professionista (iscritto albo)              |
|                                  | <b>9</b> | Non ha preferenze                                  |

**B09** (solo se B07 = 1)

**Su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo) quanto sono importanti per Lei i seguenti aspetti nella ricerca di un lavoro?**  
**(Rispondere a tutti i quesiti)**

	Per niente		Moltissimo
Prospettive di guadagno	<b>1</b>	...	<b>10</b>
Possibilità di carriera	<b>1</b>	...	<b>10</b>
Stabilità/sicurezza sul lavoro	<b>1</b>	...	<b>10</b>
Coerenza con studi fatti	<b>1</b>	...	<b>10</b>
Utilizzo delle competenze acquisite	<b>1</b>	...	<b>10</b>
Acquisizione di professionalità	<b>1</b>	...	<b>10</b>
Rispondenza ai propri interessi culturali	<b>1</b>	...	<b>10</b>
Indipendenza e autonomia sul lavoro	<b>1</b>	...	<b>10</b>
Coinvolgimento nei processi decisionali	<b>1</b>	...	<b>10</b>
Flessibilità dell'orario di lavoro	<b>1</b>	...	<b>10</b>
Tempo libero	<b>1</b>	...	<b>10</b>
Localizzazione sede di lavoro	<b>1</b>	...	<b>10</b>
Rapporti coi colleghi	<b>1</b>	...	<b>10</b>
Riconoscimento del titolo di Dottore di Ricerca	<b>1</b>	...	<b>10</b>

**FINE**

\*\*\*\*\*

\*\*\*\*\*

**Domande del Blocco C:**  
*chi non ha mai lavorato dopo il conseguimento del titolo*  
**(solo Z09 =2 )**

\*\*\*\*\*

\*\*\*\*\*

**C01**      **Lei attualmente cerca lavoro?**

- |          |    |         |
|----------|----|---------|
| <b>1</b> | Sì | (→ C02) |
| <b>2</b> | No | (→ C04) |

**C02**      **(solo se C01 = 1)**

**Quale tipo di lavoro cerca?**

- |                   |                   |          |  |
|-------------------|-------------------|----------|--|
| <b>LAVORATORE</b> | <b>DIPENDENTE</b> | <b>1</b> | Dirigente  |
|                   |                   | <b>2</b> | Direttivo/Quadro                                   |
|                   |                   | <b>3</b> | Impiegato o intermedio                             |
|                   |                   | <b>4</b> | Ricercatore o Docente universitario                |
|                   |                   | <b>5</b> | Ricercatore nel settore pubblico non universitario |
|                   |                   | <b>6</b> | Docente alle Scuole Medie (Inferiori o Superiori)  |
| <b>AUTO</b>       | <b>NOMO</b>       | <b>7</b> | Lavoro autonomo                                    |
|                   |                   | <b>8</b> | Libero professionista (iscritto albo)              |
|                   |                   | <b>9</b> | Non ha preferenze                                  |

**C03**      **(solo se C01 = 1)**

**Su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo) quanto sono importanti per Lei i seguenti aspetti nella ricerca di un lavoro?**  
**(Rispondere a tutti i quesiti)**

	Per niente		Moltissimo
Prospettive di guadagno	<b>1</b>	...	<b>10</b>
Possibilità di carriera	<b>1</b>	...	<b>10</b>
Stabilità/sicurezza sul lavoro	<b>1</b>	...	<b>10</b>
Coerenza con studi fatti	<b>1</b>	...	<b>10</b>
Utilizzo delle competenze acquisite	<b>1</b>	...	<b>10</b>
Acquisizione di professionalità	<b>1</b>	...	<b>10</b>
Rispondenza ai propri interessi culturali	<b>1</b>	...	<b>10</b>
Indipendenza e autonomia sul lavoro	<b>1</b>	...	<b>10</b>
Coinvolgimento nei processi decisionali	<b>1</b>	...	<b>10</b>
Flessibilità dell'orario di lavoro	<b>1</b>	...	<b>10</b>
Tempo libero	<b>1</b>	...	<b>10</b>
Localizzazione sede di lavoro	<b>1</b>	...	<b>10</b>
Rapporti coi colleghi	<b>1</b>	...	<b>10</b>
Riconoscimento del titolo di Dottore di Ricerca	<b>1</b>	...	<b>10</b>

**C04** (solo se C01 = 2)

**Per quale motivo non cerca lavoro?**

- 1** motivi di studio o di formazione post-dottorato
- 2** Preparazione ad un concorso pubblico
- 3** servizio civile nazionale volontario
- 4** in attesa di chiamata dopo aver superato prova / concorso / selezione
- 5** in attesa di avviare un'attività in proprio
- 6** motivi personali (casalinga/o, maternità, assistenza figli o parenti, problemi di salute, pensione, ecc.)
- 7** mancanza di opportunità di lavoro
- 8** altri motivi

**FINE**





# Appendice C

## La condizione occupazionale dei dottori di ricerca del 2008: analisi descrittiva

### *Introduzione*

In questa Appendice si procede ad un'analisi descrittiva dei dati raccolti nell'ambito della ricerca dal titolo “*Condizione attuale e prospettive occupazionali dei dottori di ricerca*” che è stata commissionata dal Comitato Nazionale per la Valutazione del Sistema Universitario (CNVSU) al Dipartimento di Statistica “G. Parenti” dell'Università degli Studi di Firenze<sup>115</sup>, a seguito di una convenzione, formalizzata in data 31 Marzo 2009.

### **C.1 L'indagine**

Il progetto di ricerca finanziato dal CNVSU ha come oggetto di indagine la condizione occupazionale dei dottori di ricerca ad uno, cinque e dieci anni dal conseguimento del titolo. I dati qui analizzati riguardano i soli dottori che hanno conseguito il titolo nell'anno 2008.

---

<sup>115</sup> Il progetto di ricerca è stato definito in seguito alla stipula di una convenzione tra il Dipartimento di Statistica “G. Parenti” ed il Ministero dell'Istruzione e della Ricerca (MIUR). Il Gruppo di Ricerca è composto dal Prof. Bruno Chiandotto (coordinatore), dalla Prof. ssa Matilde Bini, dal Dott. Bruno Bertaccini e dal Dott. Leonardo Grilli.

### *C.1.1 La popolazione*

La Popolazione obiettivo dell'indagine è costituita dai dottori di ricerca che hanno conseguito il titolo in un Ateneo italiano negli anni solari 1998, 2003 e 2008.

La popolazione di selezione<sup>116</sup> è stata individuata dal Gruppo di Ricerca chiedendo i nominativi dei dottori di ricerca appartenenti alla popolazione obiettivo agli Uffici di Supporto ai Nuclei di Valutazione di ateneo delle 71 università che hanno rilasciato titoli di dottore di ricerca, dopo che i Rettori e i Presidenti dei Nuclei di Valutazione avevano ricevuto una lettera di presentazione dell'indagine a firma del Presidente del Comitato Nazionale di Valutazione del Sistema Universitario, Prof. Luigi Biggeri.

Gli Atenei che hanno risposto e provveduto all'invio delle informazioni richieste sono stati 65. La consistenza finale della lista contenente i nominativi appartenenti alla popolazione di selezione, prima dell'avvio delle varie fasi dell'indagine, è stata di 15063 Dottori di Ricerca.

La costruzione dell'archivio informatico dei dottori di ricerca appartenenti alla popolazione di selezione e la realizzazione delle interviste è stata effettuata dalla società VALMON s.r.l.<sup>117</sup>, che ha preventivamente provveduto a sottoporre tutti i nominativi raccolti ad un processo di verifica dei recapiti presenti nell'archivio<sup>118</sup>.

### *C.1.2 La modalità di realizzazione dell'indagine*

L'indagine è stata realizzata intervistando i dottori di ricerca appartenenti alla popolazione di selezione tramite un apposito questionario predisposto dal Gruppo di Ricerca (cfr. Appendice B). Per realizzare il maggior numero di interviste, sono stati privilegiati i contatti tramite posta elettronica e telefono (fisso o cellulare), escludendo il contatto tramite posta ordinaria che notoriamente produce tassi di risposta molto bassi e quindi distorsione nelle stime statistiche (a causa dell'alta selettività dei rispondenti).

L'indagine è stata condotta da VALMON s.r.l. sotto la direzione del Gruppo di

---

<sup>116</sup> La definizione della popolazione di selezione è stata necessaria per la indisponibilità di una lista completa delle unità componenti la popolazione obiettivo.

<sup>117</sup> VALMON s.r.l. si è aggiudicata l'appalto in seguito al bando emesso dal Dipartimento di Statistica "G. Parenti" come da delibera della giunta del dipartimento in data 28/07/2009. Il contratto è stato stipulato il giorno 18 maggio 2010 con l'Università degli Studi di Firenze.

<sup>118</sup> Particolare attenzione è stata prestata alla validità degli indirizzi e-mail e dei numeri telefonici, ricavando i prefissi di telefonia fissa per i numeri che ne erano sprovvisti e ricostruendo i prefissi di telefonia cellulare che iniziavano con la cifra 0. In caso di assenza sia dell'indirizzo e-mail che del recapito telefonico, l'informazione mancante è stata integrata effettuando una ricerca ad hoc tramite internet, ma solo in relazione a nominativi a bassissimo rischio di omonimia.

Ricerca. Le interviste iniziate il 12 aprile si sono concluse il 30 luglio 2010. Alla conclusione delle quattro fasi di intervista tutti i Dottori presenti nella lista dotati di recapito di posta elettronica e/o telefonico sono stati contattati almeno una volta.

### *C.1.3 Il questionario*

Per la realizzazione della fase pilota e delle due successive fasi di intervista (CASI-web e CATI), VALMON s.r.l. si è avvalsa del software “VALMON-CATIweb”, sviluppato con tecnologie open-source per la conduzione di interviste via internet, siano esse in modalità self-interviewing (CASI) o gestite tramite intervistatore (CATI). Il questionario (Appendice B) è stato predisposto dal Gruppo di Ricerca ed è strutturato in blocchi, ognuno dei quali composto da contenuti specifici. Di seguito sono riportati, in ordine di presentazione e in maniera sintetica, gli argomenti considerati insieme ad alcuni aspetti indagati.

- Blocco Z: verifica dati di archivio (dati generali, aspetti e caratteristiche della formazione ricevuta durante il corso di dottorato).
- Blocco A: chi attualmente lavora (tempi di collocamento, caratteristiche dell'attività svolta, modalità di accesso, eventuale miglioramento della posizione professionale, grado di soddisfazione - aspetti - , partecipazione ad eventuali progetti di ricerca e numero di pubblicazioni, eventuale ricerca di un nuovo lavoro).
- Blocco B: chi attualmente non lavora ma ha lavorato dopo il conseguimento del titolo (caratteristiche ultimo lavoro svolto, eventuale miglioramento della situazione lavorativa in seguito al conseguimento del titolo, motivi dell'interruzione, eventuale ricerca o non ricerca di un nuovo lavoro - modalità e aspetti rilevanti).
- Blocco C: chi non ha mai lavorato dopo il conseguimento del titolo (eventuale ricerca o non ricerca di lavoro – tipologia, modalità e aspetti rilevanti).
- Blocco D: riepilogo attività lavorative svolte ai tempi 1, 3 5 e 10 anni (breve ricostruzione delle attività svolte ai tempi considerati).
- Blocco E: informazioni sulla famiglia di origine (titolo di studio dei genitori, stato civile dell'intervistato).

#### *C.1.4 I dati raccolti*

Il numero totale di questionari raccolti è stato di 5478 unità (Tabella C.1), 4796 interamente compilati mentre 682 parzialmente compilati. Il numero di questionari interamente compilati è prossimo al numero totale di questionari raccolti e significa che la maggior parte di coloro che hanno partecipato all'indagine lo ha fatto in maniera motivata.

Il tasso di partecipazione all'indagine è stato nel complesso pari al 36,9%. Per i dottori che hanno conseguito il titolo nel 2008 il tasso di partecipazione è stato significativamente più elevato (43,6%) mentre per i dottori che hanno conseguito il titolo nel 2003 (25,3%) e nel 1998 (18,7%) è risultato progressivamente decrescente. La differenza osservata nel tasso di partecipazione è dovuta principalmente alla qualità dei recapiti forniti dagli Atenei che si sono rivelati meno attendibili man mano che il periodo di realizzazione dell'indagine si allontanava da quello di conseguimento del titolo.

*Tabella C.1. Popolazione di selezione, numero di questionari e tasso di partecipazione (%) dell'indagine per anno di conseguimento del titolo di dottorato.*

<b>Anno</b>	<b>Popolazione di selezione</b>	<b>Numero di questionari</b>	<b>Tasso di partecipazione (%)</b>
1998	673	126	18,7
2003	4462	1129	25,3
2008	9696	4223	43,6
Totale	14831	5478	36,9

Una volta esaurite tutte le fasi di realizzazione dell'indagine, la consistenza effettiva della lista degli appartenenti alla popolazione di selezione è scesa a 14831 unità. Questa diminuzione<sup>119</sup> è dovuta principalmente all'intervento dei soggetti intervistati che hanno provveduto a correggere le informazioni anagrafiche fornite dagli Atenei<sup>120</sup>

#### *C.1.5 Il confronto di struttura*

Il confronto di struttura tra la popolazione (definita dai nominativi inviati dagli atenei) ed il campione è stato fatto per area disciplinare del dottorato, distintamente per anno di

---

<sup>119</sup> Solo in rare situazioni le correzioni apportate dai dottori in sede di intervista hanno portato ad ottenere interviste a persone inizialmente non contattabili.

<sup>120</sup> Principalmente si è trattato di variazioni dell'anno di conseguimento del titolo che hanno prodotto l'esclusione dei soggetti intervistati dalla piattaforma di somministrazione del questionario.

conseguimento del titolo (1998, 2003 e 2008)<sup>121</sup>. Di seguito è riportato il solo confronto relativo ai dottori che hanno conseguito il titolo nell'anno 2008 (Tabella C.2).

Il tasso di risposta complessivo supera il 40% con un minimo del 30,8% per l'area "Scienze Pedagogiche" ed un massimo del 50,4% per l'area "Scienze Agrarie". Nel complesso il campione relativo al 2008 consente un'analisi per area disciplinare poiché il numero di dottori intervistati è significativo per ogni area ed i tassi di risposta non sono molto diversi tra loro.

*Tabella C.2. Dottori di ricerca del 2008 contattabili vs intervistati per area disciplinare del corso di dottorato.*

	Popolazione		Campione		Tasso
	N	%	N	%	
Ingegneria civile e Architettura	569	6,1	231	6,0	40,6
Ingegneria dell'informazione	317	3,4	136	3,5	42,9
Ingegneria industriale	483	5,2	204	5,3	42,2
Scienze agrarie	345	3,7	174	4,5	50,4
Scienze biologiche	1038	11,2	504	13,1	48,6
Scienze chimiche	619	6,7	231	6,0	37,3
Scienze della terra	445	4,8	179	4,7	40,2
Scienze dell'antichità e filologico-letterarie	606	6,5	297	7,7	49,0
Scienze delle attività motorie e discipline sportive	44	0,5	20	0,5	45,5
Scienze economiche (aziendali)	190	2,0	77	2,0	40,5
Scienze economiche (socio-politiche)	349	3,8	132	3,4	37,8
Scienze fisiche	522	5,6	234	6,1	44,8
Scienze giuridiche	688	7,4	241	6,3	35,0
Scienze informatiche	49	0,5	21	0,5	42,9
Scienze matematiche	153	1,6	67	1,7	43,8
Scienze mediche	1361	14,7	486	12,6	35,7
Scienze pedagogiche	130	1,4	40	1,0	30,8
Scienze politiche e sociali	289	3,1	127	3,3	43,9
Scienze psicologiche, geografiche e demotnoantropologiche	254	2,7	106	2,8	41,7
Scienze statistiche	56	0,6	20	0,5	35,7
Scienze storiche e filosofiche	456	4,9	183	4,8	40,1
Scienze storico-artistiche	184	2,0	81	2,1	44,0
Scienze veterinarie	128	1,4	57	1,5	44,5
<b>Totale</b>	<b>9275</b>	<b>100,0</b>	<b>3848</b>	<b>100,0</b>	<b>41,5</b>

<sup>121</sup> I valori delle frequenze sui quali è stato fatto il confronto sono leggermente diversi da quelli rispetto a quelli della Tabella 3.4 in quanto per alcuni dottori non è stato possibile identificare l'area di appartenenza del corso di dottorato. Tale differenza non può comunque ritenersi significativa.

## C.2 La condizione occupazionale dei dottori di ricerca del 2008

In questo capitolo si presentano i risultati dell'indagine condotta sui dottori di ricerca che hanno conseguito il titolo nel 2008. La sequenza nella presentazione e il commento sui dati rispecchia quella prevista nella scheda di rilevazione utilizzata (riportata integralmente in Appendice B).

### C.2.1 La formazione ricevuta

#### C.2.1.1 Le motivazioni dell'iscrizione al corso di dottorato

Il motivo principale per cui i dottori di ricerca hanno deciso di iscriversi al corso di dottorato in cui hanno conseguito il titolo (cfr. Tabella C.Z01) è stato *per fare attività di ricerca* (41,4%). Tale motivazione è più diffusa tra i dottori appartenenti alle aree disciplinari delle *scienze informatiche* (61,9%) e delle *scienze fisiche* (58,8%) e piuttosto limitata tra i dottori delle *scienze statistiche* (16,7%) e delle *scienze giuridiche* (19,4%) (cfr. Tabella C.Z01a e Tabella C.Z01b). Invece, per quanto riguarda il genere, sono interessati all'attività di ricerca le *femmine* (43,8%) più dei *maschi* (38,7%).

Più di un quinto dei dottori è stato spinto ad intraprendere gli studi *per l'interesse nei confronti di una futura carriera universitaria* (20,3%) che sembra quindi essere ancora attraente, in particolare per i dottori delle *scienze economiche (aziendali)* (36,8%), delle *scienze giuridiche* (35,1%) e delle *scienze dell'antichità e filologico-letterarie* (32,2%) e, per quanto riguarda il genere, per i *maschi* (22,7%) più che per le *femmine* (18,2%).

La decisione di iscriversi è sufficientemente motivata anche dalla possibilità di *completare/arricchire la formazione* (19,5%), soprattutto per i dottori delle *scienze mediche* (32,1%).

Pochi dottori hanno deciso di iscriversi ad un corso di dottorato *per conseguire un titolo dal più alto potere professionalizzante* (10,4%), a conferma che il titolo di dottore di ricerca sembra essere poco desiderabile per chi vuole accrescere la propria formazione a fini professionali. Il valore professionalizzante del titolo è apprezzato soprattutto dai dottori in *scienze delle attività motorie e delle discipline sportive* (25,0%) e da quelli nelle *scienze economiche (aziendali)* (17,1%) ma quasi per niente

dai dottori in *scienze fisiche* (4,1%), *scienze matematiche* (4,4%) e *scienze storiche e filosofiche* (4,7%).

Coloro che sono stati spinti ad intraprendere questa scelta solo dall'*interesse per lo studio* rappresentano invece una minoranza (6,4%). Questi sono soprattutto dottori in *scienze matematiche* (19,1%), *scienze statistiche* (16,7%) e *scienze giuridiche* (13,2%).

In generale, il grado di associazione tra il motivo dell'iscrizione e l'area disciplinare del corso di dottorato di ricerca è piuttosto moderato (V di Cramer: 0,154; gdl: 110;  $p < 0,001$ ) e limitato alle eccezioni descritte. Il motivo dell'iscrizione si differenzia appena per genere (V di Cramer: 0,066; gdl: 5;  $p = 0,003$ ).

#### C.2.1.2 La valutazione dell'esperienza formativa

Il livello di soddisfazione riguardo la formazione ricevuta è stato valutato rispetto ad una serie di aspetti ritenuti rilevanti, misurati su una scala ordinale a dieci modalità di risposta<sup>122</sup> (cfr. Tabelle C.Z03.1-C.Z03.5). La distribuzione di frequenze degli aspetti considerati presenta un certo grado di asimmetria<sup>123</sup> negativa ed indica che le modalità più frequenti sono quelle centrali ed elevate.

L'aspetto che ha ricevuto la valutazione migliore è stato (cfr. Tabella C.Z03c) la *qualità scientifica di professori/ricercatori* (mediana: 8; asimmetria: -1,10).

Lo *sviluppo/addestramento della capacità di fare ricerca* (mediana: 7; asimmetria: -0,66), la *qualità della didattica del personale docente* (mediana: 7; asimmetria: -0,68) e la *possibilità di pubblicare (articoli, volumi, ...)* (mediana: 7; asimmetria: -0,80) hanno ricevuto una valutazione buona ma inferiore.

Il *contenuto degli insegnamenti* ha ricevuto invece una valutazione appena sufficiente (mediana: 6; asimmetria: -0,46).

La scelta di iscriversi ad un corso di dottorato non è sempre stata fatta in maniera convinta (cfr. Tabella C.Z04). Infatti, solo poco più della metà dei dottori, *se potesse tornare indietro, si iscriverebbe allo stesso corso nella stessa sede universitaria* (52,9%). In generale, ripeterebbero la stessa scelta i maschi (57,4%) più delle femmine (48,9%). Invece, per quanto riguarda le aree disciplinari, i più convinti sono i dottori in

<sup>122</sup> A ciascuna modalità è stato assegnato un punteggio al quale corrispondono i valori 1= per niente e 10 = moltissimo.

<sup>123</sup> Per misurare il grado di asimmetria è stato utilizzato l'indice di skewness, ottenuto  $\gamma_1 = m_3/m_2^{3/2}$ , dove  $m_2$  ed  $m_3$  sono rispettivamente il secondo e terzo momento centrale.

*ingegneria industriale* (63,2%), *scienze giuridiche* (63,1%), *scienze pedagogiche* (62,5%) ed *ingegneria dell'informazione* (61,4%) mentre quelli meno convinti sono i dottori in *scienze politiche e sociali* (45,1%), *scienze biologiche* (45,7%) e *scienze economiche (socio-politiche)* (46,2%).

Gli altri dottori che si iscriverebbero di nuovo ad un corso di dottorato di ricerca lo farebbero invece *allo stesso corso ma in un'altra sede universitaria* (18,8%) o *ad un altro corso* (11,7%).

Un sesto dei dottori di ricerca ha dichiarato invece che *non si iscriverebbe di nuovo ad un corso di dottorato* (16,7%). Questi sono soprattutto i dottori in *scienze agrarie* (22,4%), *scienze chimiche* (21,0%), *scienze mediche* (20,9%) e *scienze biologiche* (20,1%) (cfr. Tabella C.Z04a e Tabella C.Z04b). Invece, per quanto riguarda il genere, sono *femmine* (18,8%) più che *maschi* (14,1%).

In definitiva, la decisione di iscriversi di nuovo presenta un grado di associazione piuttosto moderato con l'area disciplinare del corso di dottorato (V di Cramer: 0,131; gdl: 66;  $p < 0,001$ ) e con il genere (V di Cramer: 0,102; gdl: 5;  $p < 0,001$ ), limitato alle sole eccezioni descritte.

#### C.2.1.3 L'esperienza lavorativa durante gli studi

La maggior parte dei dottori (cfr. Tabella C.Z06) ha svolto qualche attività lavorativa durante gli studi (58,4%). Per più della metà di loro si è trattato di un'*attività lavorativa attinente all'attività di ricerca del corso* (34,3%) mentre per più di un terzo si è trattato di un'*attività non attinente* (24,1%).

I dottori che hanno condiviso l'esperienza di studio con quella lavorativa sono stati soprattutto quelli appartenenti alle aree delle *scienze economiche (aziendali)* (84,2%), *scienze politiche e sociali* (78,0%), *scienze economiche (socio-politiche)* (76,0%), *ingegneria civile e architettura* (73,0%) e *scienze informatiche* (71,4%) (cfr. Tabella C.Z06a e Tabella C.Z06b).

Invece, i dottori che si sono dedicati esclusivamente agli studi sono in prevalenza quelli delle *scienze biologiche* (55,2%), *scienze chimiche* (53,2%), *scienze fisiche* (53,1%), *scienze veterinarie* (48,2%) e *scienze matematiche* (45,6%).

In generale, svolgere una qualche attività lavorativa presenta un grado di associazione debole con l'area disciplinare del corso di dottorato (V di Cramer: 0,192;



gdl: 44;  $p < 0,001$ ) e nulla con il genere (V di Cramer: 0,029; gdl: 2;  $p = 0,169$ ).

Al momento di conseguimento del titolo, la percentuale di chi ha dichiarato di lavorare è risultata significativamente inferiore (46,7%), a dimostrazione che l'iscrizione al dottorato ha spinto una parte dei dottori a dedicarsi completamente agli studi (cfr. Tabella C.Z07).

#### C.2.1.4 *Il periodo di studi*

Durante il periodo di studi, la maggioranza dei dottori (cfr. Tabella C.Z02) è stata titolare di *un posto con borsa di studio finanziata dall'università/MIUR* (57,9%) o *da altri enti pubblici o privati* (11,7%). Quasi un quinto dei dottori ha avuto un *posto senza borsa studio e non coperto da assegno di ricerca* (18,0%) mentre gli altri, pur non avendo la borsa, hanno ricevuto almeno *un assegno di ricerca* (6,1%).

La percentuale di coloro che sono stati titolari di una borsa di studio sale in maniera significativa tra coloro che *non hanno svolto alcuna attività lavorativa durante il corso* (72,4%) e scende tra coloro che hanno svolto un'*attività lavorativa attinente* (47,1%) o *non attinente* (48,2%) *all'attività di ricerca del corso*.

La maggioranza dei dottori *non ha compiuto studi all'estero* (60,4%) durante il corso di dottorato (cfr. Tabella C.Z05). In particolare, questi sono soprattutto i dottori nelle *scienze storico-artistiche* (76,9%), *scienze mediche* (76,8%), *scienze biologiche* (67,7%) e *scienze pedagogiche* (67,6%). Invece, hanno avuto un'esperienza di studi all'estero soprattutto i dottori nelle *scienze statistiche* (69,6%), *scienze economiche (socio-politiche)* (55,8%), *scienze politiche e sociali* (53,2%) e *scienze psicologiche, geografiche e demotnoantropologiche* (52,4%) (cfr. Tabella C.Z05a e Tabella C.Z05b).

#### C.2.1.5 *La condizione occupazionale attuale*

Ad un anno dal conseguimento del titolo, la maggior parte dei dottori (cfr. Tabella C.Z08) ha dichiarato di svolgere un'attività lavorativa retribuita (83,8%). In particolare, si tratta dei dottori appartenenti all'area delle *scienze fisiche* (92,7%), *scienze statistiche* (91,7%), *ingegneria dell'informazione* (91,1%), *scienze politiche e sociali* (90,6%), *ingegneria industriale* (90,3%), *scienze informatiche* (90,0%) e *scienze economiche (aziendali)* (89,3%) (cfr. Tabella C.Z07a e Tabella C.Z07b). Invece, per quanto riguarda il genere, sono *maschi* (86,4%) più

che le *femmine* (81,5%).

I dottori ancora alla *ricerca di un lavoro* (5,7%) o che si trovano in un'*altra condizione di inattività (non lavora e non cerca lavoro)* (0,5%) sono invece in pochi, in prevalenza appartenenti all'area delle *scienze storiche e filosofiche* (13,6%) e *scienze dell'antichità e filologico-letterarie* (12,1%).

Una quota quasi trascurabile dei dottori ha dichiarato invece di svolgere ancora un'attività di formazione, *post-dottorato retribuita* (1,6%), *post-dottorato non retribuita* (1,3%) o si trovavano in un *periodo di studio in preparazione a concorsi pubblici/esame* (1,0%). Nel complesso, questi sono soprattutto i dottori nelle *scienze storico-artistiche* (11,3%) e *scienze giuridiche* (8,6%).

La condizione occupazionale prevalente dichiarata è il *lavoro con assegno di ricerca, borsa post-doc o altra borsa* (35,4%), tipica per un dottore di ricerca. I ricercatori sono soprattutto dottori nelle aree delle *scienze matematiche* (61,2%), delle *scienze fisiche* (57,5%) e delle *scienze agrarie* (50,3%), indifferentemente maschi (35,3%) o femmine (35,4%) (cfr. Tabella C.Z08a e Tabella C.Z08b).

Il *lavoro a tempo indeterminato/autonomo effettivo* riguarda invece poco meno di un terzo dei dottori di ricerca (26,9%). I lavoratori con una forma contrattuale stabile sono in prevalenza maschi (31,3%) rispetto a femmine (23,3%) mentre per quanto riguarda le aree disciplinari appartengono soprattutto alle *scienze economiche (aziendali)* (48,0%), *scienze pedagogiche* (43,6%), *ingegneria industriale* (43,6%) e *scienze giuridiche* (42,2%).

Il *lavoro con altra forma contrattuale a tempo determinato (contratto formazione lavoro, collaborazione, altro contratto atipico, ecc)* riguarda una percentuale di dottori significativamente inferiore (21,5%), in particolare quelli delle *scienze informatiche* (40,0%) e delle *scienze dell'antichità e filologico-letterarie* (34,2%).

In generale, il grado di associazione tra la condizione occupazionale è piuttosto moderato con l'area disciplinare del corso di dottorato di ricerca (V di Cramer: 0,130; gdl: 176;  $p < 0,001$ ) e con genere (V di Cramer: 0,112; gdl: 8;  $p < 0,001$ ).

La condizione occupazionale attuale non è indipendente dalla condizione al momento del conseguimento del titolo di dottore. Infatti, tra coloro che già lavoravano al momento del conseguimento del titolo di dottore, la percentuale di chi ha dichiarato

di svolgere un *lavoro a tempo indeterminato/autonomo effettivo* è significativamente maggiore (41,8%) mentre la percentuale di chi ha dichiarato di svolgere un *lavoro con assegno di ricerca, borsa post-doc o altra borsa* è significativamente maggiore tra chi non lavorava (44,6%).

### C.3 Chi attualmente lavora

#### C.3.1 Il lavoro attuale

Più di due terzi dei dottori (cfr. Tabella C.A02) hanno dichiarato di fare *lo stesso lavoro che svolgevano al momento del conseguimento del titolo* (69,3%). Per la maggior parte di questi il *conseguimento del titolo di dottore di ricerca non ha comportato un miglioramento del lavoro attuale* (62,2%). Un quadro poco confortante ma che potrebbe modificarsi in futuro, considerato il poco tempo trascorso dal conseguimento del titolo e l'attuale rigidità del mercato del lavoro.

I dottori che hanno dichiarato di aver avuto *un miglioramento del lavoro attuale dopo il conseguimento del titolo* rappresentano comunque una percentuale non trascurabile (450 su 1192, pari al 37,8%). Tale percentuale è significativamente maggiore per chi ha un *lavoro con assegno di ricerca, borsa post-doc o altra borsa* (52,9%) e minore per chi ha un *lavoro a tempo indeterminato/autonomo effettivo* (31,3%), a dimostrazione che gli effetti del dottorato sull'attività lavorativa sono più evidenti, almeno nella prima fase dopo il conseguimento del titolo, proprio per chi si dedica all'attività di ricerca. Per quanto riguarda le aree disciplinari, il titolo di dottore ha comportato un miglioramento solo per i dottori delle *scienze informatiche* (57,1%), delle *scienze chimiche* (49,1%) e delle *scienze fisiche* (46,2%) e, per quanto riguarda il genere, appena di più per i *maschi* (40,7%) rispetto alle *femmine* (35,1%).

Il miglioramento del lavoro attuale non presenta un'associazione significativa con l'area disciplinare (V di Cramer: 0,154; gdl: 22; p = 0,229) mentre la relazione con il genere è molto debole (V di Cramer: 0,058; gdl: 1; p = 0,046).

A chi ha dichiarato di aver avuto un miglioramento nel proprio lavoro, è stato chiesto di indicare quali fossero gli aspetti migliorati, tra quelli loro elencati<sup>124</sup> (cfr.

---

<sup>124</sup> Per ciascun aspetto, le modalità di risposta erano “sì”, se migliorato e “no”, se non migliorato.

Tabelle C.A06.1-C.A06.16).

In generale, l'effetto del titolo di dottore è risultato piuttosto moderato se valutato in relazione al *miglioramento generale rispetto alle aspettative iniziali* (57,5%) e al *miglioramento generale rispetto all'attività lavorativa ideale* (44,6%).

In particolare, i dottori hanno riscontrato un miglioramento significativo soprattutto nell'*acquisizione di professionalità* (91,2%) e nell'*utilizzo delle competenze acquisite* (90,4%) e, sebbene in maniera minore, anche nella *rispondenza ai propri interessi culturali* (85,0%) e nella *coerenza con gli studi fatti* (80,9%). Per questi dottori il titolo sembra aver avuto effetti decisamente positivi nello svolgimento dell'attività lavorativa.

Il titolo di dottore ha contribuito anche ad incrementare le *possibilità di carriera* (70,7%), l'*indipendenza e autonomia sul lavoro* (66,2%) ed il *coinvolgimento nei processi decisionali* (56,2%).

Non si registrano effetti positivi sulla *stabilità/sicurezza sul lavoro* (21,1%), sulla *localizzazione della sede di lavoro* (18,1%) e sul *tempo libero* (12,2%).

### C.3.2 Tempi e modalità di ingresso nel mondo del lavoro

Il tempo necessario per trovare il primo lavoro iniziato dopo il conseguimento del titolo di dottore di ricerca è stato generalmente limitato. Quasi la metà dei dottori ha impiegato meno di un mese (47,8%), più di tre quarti ci sono riusciti entro sei mesi (78,2%) mentre quasi tutti avevano un lavoro dopo un anno (93,2%) o dopo un anno e mezzo (97,4%). Nel complesso, la distribuzione dei tempi di inserimento nel mondo del lavoro è asimmetrica a destra e leptocurtica (mediana: 1; asimmetria: 1,91; curtosi: 2,85) e significa che la maggior parte delle frequenze è addensata nei primi mesi dopo il conseguimento del titolo mentre poche risposte si trovano su valori più elevati e distanti da queste (cfr. Tabella C.A07c).

I più rapidi ad inserirsi nel mondo del lavoro sono stati i dottori delle *scienze informatiche* (mediana: 0; asimmetria: 3,32; curtosi: 11,00), *ingegneria dell'informazione* (mediana: 0; asimmetria: 2,70; curtosi: 7,29), *ingegneria industriale* (mediana: 0; asimmetria: 2,72; curtosi: 8,35) e delle *scienze fisiche* (mediana: 0; asimmetria: 1,83; curtosi: 2,73). Tempi di inserimento più lunghi si registrano invece per i dottori nelle aree delle *scienze storiche e filosofiche* (mediana: 5; asimmetria: 0,58;

curtosi: -0,59), delle *scienze statistiche* (mediana: 5; asimmetria: 1,77; curtosi: 3,64), delle *scienze giuridiche* (mediana: 3,5; asimmetria: 1,20; curtosi: 1,39), delle *scienze dell'antichità e filologico-letterarie* (mediana: 3; asimmetria: 1,22; curtosi: 0,85) e delle *scienze storico-artistiche* (mediana: 3; asimmetria: 1,40; curtosi: 1,37). Invece, per quanto riguarda il genere, i *maschi* hanno impiegato meno tempo (mediana: 0; asimmetria: 1,82; curtosi: 3,00) rispetto alle *femmine* (mediana: 2; asimmetria: 1,69; curtosi: 2,64).

Il canale prevalentemente utilizzato per trovare lavoro (cfr. Tabella C.A10) è stato la *domanda o partecipazione a concorsi pubblici* (33,0%), modalità tipica necessaria per l'accesso ad un'attività in enti pubblici di ricerca. Questa percentuale è risultata naturalmente maggiore per chi ha dichiarato di svolgere un *lavoro con assegno di ricerca, borsa post-doc o altra borsa* (40,4%) rispetto a chi svolge un *lavoro con altra forma contrattuale a tempo determinato* (18,0%). Il dottorato ha avuto un ruolo non trascurabile anche come canale per trovare lavoro, considerato che una percentuale significativa di dottori ha avuto *contatti con datori di lavoro su segnalazione di università/docenti* (17,1%).

I *contatti con datori di lavoro su iniziativa personale o altra iniziativa* hanno riguardato invece un sesto dei dottori (16,6%) mentre gli altri si sono serviti delle modalità più diverse, *siti internet e/o motori di ricerca* (7,5%), *risposta a inviti a partecipare a colloqui da parte di aziende* (1,5%) ed *altre* ancora per un totale che supera un quinto del totale (20,4%).

### C.3.3 La tipologia del lavoro attuale

La quasi totalità dei dottori ha un *lavoro dipendente (inclusi contratti di collaborazione)* (87,2%) (cfr. Tabella C.A11). I *lavoratori autonomi* rappresentano invece una quota marginale (12,8%).

In particolare, i lavoratori dipendenti appartengono soprattutto all'ambito disciplinare delle *scienze matematiche* (96,6%), delle *scienze chimiche* (95,6%), delle *scienze fisiche* (95,2%) e delle *scienze biologiche* (94,2%) mentre i lavoratori autonomi sono soprattutto quelli degli ambiti tipici della libera professione, *scienze giuridiche* (39,5%), *ingegneria civile e architettura* (32,8%) e *scienze economiche (aziendali)* (27,7%) (cfr. Tabella C.CA11a e Tabella C.A11b). Non si osservano differenze

significative dovute al genere (V di Cramer: 0,059; gdl: 1;  $p < 0,032$ ).

I lavoratori dipendenti hanno in prevalenza (cfr. Tabella C.A.12) la forma contrattuale dell'*assegno di ricerca o la borsa di studio* (37,3%) anche se più di un quarto dei dottori ha un impiego a *tempo pieno e indeterminato* (27,9%) mentre quasi un sesto ha un impiego a *tempo pieno, a termine* (14,9%). Il rimanente si divide tra le altre forme contrattuali cosiddette atipiche, *collaborazione a progetto* (7,5%), *collaborazione coordinata e continuativa* (4,6%), *part-time a termine* (2,3%), *collaborazione occasionale* (1,1%), *prestazione d'opera* (0,8%) ed *altro* (1,5%). Una quota trascurabile di dottori svolge un'attività non regolata da contratto (0,6%).

I lavoratori autonomi sono in prevalenza *liberi professionisti iscritti all'albo* (57,6%) o *non iscritti all'albo* (36,4%) mentre gli imprenditori costituiscono una quota marginale (6,1%).

#### C.3.4 Le mansioni di ricerca ed innovazione

I dottori con lavoro autonomo o dipendente che non sono ricercatori, *non svolgono mansioni di ricerca e sviluppo* (43,7%) o, se la svolgono, questa *non costituisce l'attività prevalente* (42,4%) mentre la ricerca e lo sviluppo costituiscono l'*attività prevalente* solo in casi limitati (13,9%) (cfr. Tabella C.A16).

I dottori che svolgono attività di ricerca e sviluppo in modo prevalente sono quelli appartenenti all'ambito disciplinare di *ingegneria industriale* (31,4%), *ingegneria dell'informazione* (26,3%) e, in misura minore, *scienze chimiche* (21,8%) e *scienze della terra* (21,2%) (cfr. Tabella C.A16a e Tabella C.A16b).

L'attività di innovazione è invece più diffusa (45,6%). In particolare, si tratta di *innovazione di processo e di prodotto* (19,4%) o *innovazione di processo* (17,1%) mentre l'*innovazione di prodotto* è più rara (9,2%) (cfr. Tabella C.A17).

L'innovazione di processo e di prodotto riguarda soprattutto i dottori in *ingegneria industriale* (35,3%), *scienze fisiche* (34,4%) e *scienze chimiche* (29,1%) (cfr. Tabella C.A17a e Tabella C.A17b). L'innovazione di processo riguarda soprattutto i dottori in *scienze economiche (aziendali)* (38,9%) e *scienze biologiche* (24,6%). L'innovazione di prodotto riguarda invece i dottori *ingegneria dell'informazione* (42,1%) e, in misura molto minore, i dottori in *scienze fisiche* (21,9%).

### C.3.5 Il profilo giuridico del lavoro dipendente

I dottori che hanno un lavoro dipendente appartengono in prevalenza (cfr. Tabella C.A14) all'area della ricerca (59,1%), nelle sue diverse forme giuridiche. Questi sono soprattutto i dottori delle *scienze fisiche* (77,2%), delle *scienze agrarie* (76,1%), delle *scienze veterinarie* (73,7%), delle *scienze biologiche* (69,3%), delle *scienze informatiche* (66,7%) e delle *scienze psicologiche, geografiche e demoetnoantropologiche* (66,7%). Invece, per quanto riguarda le differenze di genere, i *maschi* prevalgono appena (60,4%) rispetto alle *femmine* (57,9%) (cfr. Tabella C.A14a e Tabella C.A14b).

La forma contrattuale più rappresentata è quella di *assegnista/borsista* (39,5%) mentre i dottori già inseriti a vario titolo in modo più stabile in ambito universitario rappresentano una quota significativamente minore ma non trascurabile (9,3%). In particolare, si tratta di *ricercatori universitari a tempo determinato* (5,4%), *ricercatori universitari di ruolo* (3,5%) e, sebbene rappresentino una quota marginale, anche di *professori universitari associati* (0,3%) e *professori universitari ordinari* (0,1%).

Tra le altre forme contrattuali di lavoro dipendente, quasi un quarto sono *dirigenti* (5,0%), *direttivi/quadri* (3,5%) e *impiegati o intermedi* (15,6%). In prevalenza si tratta di dottori appartenenti all'ambito delle *scienze economiche (aziendali)* (38,3%), delle *scienze giuridiche* (38,1%), delle *scienze mediche* (34,3%), delle *scienze informatiche* (33,3%) e dell'*ingegneria dell'informazione* (33,0%) e, per quanto riguarda il genere, di *maschi* (26,7%) più che di *femmine* (21,8%).

Il resto dei lavoratori dipendenti lavora in ambito scolastico, e sono docenti alle scuole medie inferiori (2,2%) o docenti alle scuole medie superiori (4,6%). Questi sono soprattutto dottori nell'ambito delle *scienze dell'antichità e filologico-letterarie* (27,3%), delle *scienze storiche e filosofiche* (21,3%), delle *scienze pedagogiche* (20,7%) e delle *scienze storico-artistiche* (17,6%).

In generale, il grado di associazione tra il profilo giuridico del lavoro dipendente è piuttosto debole con l'area disciplinare del corso di dottorato di ricerca (V di Cramer: 0,180; gdl: 308;  $p < 0,001$ ) e con il genere (V di Cramer: 0,150; gdl: 14;  $p < 0,001$ ).

### C.3.6 Le caratteristiche dell'attività di ricerca

La maggior parte dei dottori che lavorano in abito universitario (cfr. Tabella C.A18) *presta servizio nello stesso ateneo in cui ha svolto il dottorato* (64,4%) mentre gli altri sono *in un altro ateneo italiano* (20,9%) o *in un ateneo estero* (14,7%). La percentuale di quanti prestano servizio nello stesso ateneo è maggiore tra gli *assegnisti/borsisti* (70,0%) e minore tra i *ricercatori universitari a tempo determinato* (32,0%) che, in prevalenza, si trovano a fare ricerca all'estero (54,9%) mentre i ricercatori universitari di ruolo si trovano soprattutto nello stesso ateneo in cui hanno conseguito il titolo di dottore.

I dottori che svolgono un'attività di ricerca (cfr. Tabella C.A19), lo fanno prevalentemente in un *ambito pienamente coerente con la formazione acquisita nel dottorato di ricerca* (67,3%) o *in un ambito parzialmente coerente* (28,2%). L'attività di ricerca è *pienamente coerente* soprattutto per i *ricercatori universitari di ruolo* (78,8%) e per gli *assegnisti/borsisti* (70,1%) mentre lo è molto meno per i *ricercatori del settore privato* (37,5%). Invece, tra gli ambiti disciplinari, quelli in cui l'attività di ricerca è pienamente coerente sono quelli delle *scienze storico-artistiche* (94,1%), delle *scienze dell'antichità e filologico-letterarie* (81,4%), delle *scienze storiche e filosofiche* (77,1%) e delle *scienze politiche e sociali* (76,7%) (cfr. Tabella C.A19a e Tabella C.A19b).

Dopo il conseguimento del titolo di dottore di ricerca, la maggior parte dei dottori (cfr. Tabelle C.A20c-C.A23c) *non ha pubblicato lavori su riviste scientifiche a diffusione nazionale* (52,1%) mentre la percentuale di coloro che *non hanno pubblicato lavori su riviste scientifiche a diffusione internazionale* è stata notevolmente inferiore (16,2%), probabilmente a causa del minor prestigio e della minore diffusione delle prime rispetto alle ultime. In generale, il numero medio di *lavori pubblicati* è risultato più di uno e mezzo (1,61) *su riviste scientifiche a diffusione nazionale* e di quasi quattro (3,92) *su riviste scientifiche a diffusione internazionale*. In particolare, *sulle riviste scientifiche a diffusione nazionale i più attivi sono stati i ricercatori universitari di ruolo* (2,55) mentre *sulle riviste scientifiche a diffusione internazionale i ricercatori del settore pubblico non universitario* (4,81) ed i *ricercatori universitari di ruolo* (4,60). Svolgere un'attività di ricerca coerente con la formazione acquisita nel corso di dottorato contribuisce ad incrementare leggermente solo il numero medio di pubblicazioni *su riviste scientifiche a diffusione internazionale* (4,32) ma non sulle



*riviste scientifiche a diffusione nazionale* (1,71).

La maggior parte dei dottori *non ha pubblicato monografie da editore nazionale* (81,4%) o *da editore internazionale* (81,4%) mentre più di due terzi hanno partecipato a *progetti di ricerca nazionali* (69,7%) e poco più della metà a *progetti di ricerca nazionali* (48,0%).

### C.3.7 *Le caratteristiche del contesto di lavoro*

I dottori che non lavorano in ambito universitario, della scuola o negli enti pubblici di ricerca (cfr. Tabella C.A24), sono impiegati soprattutto negli *enti pubblici* (44,3%) o nelle *aziende private italiane (anche se in filiale all'estero)* (31,9%) mentre gli altri si dividono tra le *aziende private straniere (anche se in filiale in Italia)* (11,5%) e le *aziende a partecipazione pubblica* (9,0%) mentre in pochi sono nelle *organizzazioni no-profit* (3,4%).

Negli enti pubblici lavorano soprattutto i *dirigenti* (78,9%) ed i *direttivi/quadri* (53,3%) mentre nelle aziende private italiane (anche se in filiale all'estero) lavorano prevalentemente gli *impiegati o intermedi* (43,1%).

I dottori lavorano prevalentemente in enti (aziende o in organizzazioni no-profit) di grandi dimensioni, *con più di 250 dipendenti* (45,4%), mentre gli altri si dividono abbastanza equamente tra le altre tipologie dimensionali, *fino a 10 dipendenti* (17,1%), *da 11 a 50 dipendenti* (17,3%) e *da 51 a 250 dipendenti* (20,2%) (cfr. Tabella C.A25). Le aziende di grandi dimensioni sono prevalentemente quella a partecipazione pubblica (79,4%) o le aziende private straniere (anche se in filiale in Italia) (69,5%).

Il ramo di attività economica più diffuso è quello della *sanità* che impiega quasi un quarto dei dottori di ricerca (23,6%). Gli altri sono impiegati nell'*istruzione, ricerca e sviluppo* (13,0%) o nella *Pubblica Amministrazione, Forze Armate* (12,8%) mentre una quota minore ma non trascurabile sono impiegati nella *chimica* (6,9%), nell'*informatica, elaborazione dati* (4,7%) o nel *credito e assicurazioni* (4,1%) (cfr. Tabella C.A26).

### C.3.8 *L'utilità sul lavoro della formazione ricevuta durante il corso di dottorato*

La formazione ricevuta durante il corso di dottorato è stata giudicata (cfr. Tabella C.A29) *fondamentale per le mansioni svolte* (39,6%) o *utile per il proprio lavoro anche*

se non ne fa un uso specifico (38,7%). Invece, chi ne fa un uso piuttosto limitato (12,4%) o non ne fa alcun uso (9,3%) rappresenta una minoranza non trascurabile.

La formazione ricevuta durante il corso di dottorato è ritenuta *fondamentale per le mansioni svolte* soprattutto per chi lavora in ambito universitario, *professore ordinario* (75,0%), *professore associato* (75,0%), *ricercatore universitario di ruolo* (67,3%) e *assegnista/borsista* (63,8%). Invece, non ne fanno alcun uso prevalentemente gli *impiegati o intermedi* (26,3%), i *docenti delle scuole medie inferiori* (26,2%) ed i *direttivi/quadri* (16,5%).

Per quanto riguarda gli ambiti disciplinari, la formazione ricevuta durante il corso di dottorato è ritenuta fondamentale (cfr. Tabella C.A29a e Tabella C.A29b) per i dottori delle *scienze informatiche* (58,8%), delle *scienze matematiche* (58,2%), delle *scienze agrarie* (50,4%), delle *scienze statistiche* (50,0%), dell'*ingegneria industriale* (49,7%) e delle *scienze fisiche* (49,5%).

Invece, per quanto riguarda il genere, la formazione ricevuta durante il corso di dottorato è ritenuta fondamentale dai *maschi* (42,9%) più che dalle *femmine* (39,6%).

### C.3.9 Il livello di soddisfazione per il lavoro attuale

Il livello di soddisfazione per il lavoro attuale è stato valutato rispetto ad una serie di aspetti ritenuti rilevanti misurati su una scala ordinale a dieci modalità di risposta<sup>125</sup> (cfr. Tabella C.A30.1- C.A30.18). Il livello di soddisfazione per il lavoro attuale è stato giudicato (cfr. Tabella C.A30c) discreto *nel suo complesso* (mediana: 7; asimmetria: -0,81; curtosi: 0,93), *rispetto alle mansioni svolte* (mediana: 7; asimmetria: -0,92; curtosi: 0,78) e *rispetto alle aspettative iniziali* (mediana: 7; asimmetria: -0,67; curtosi: -0,12) ma appena sufficiente *rispetto al titolo di dottore di ricerca* (mediana: 6; asimmetria: -0,44; curtosi: -0,87).

Gli aspetti che hanno ricevuto una valutazione migliore sono quelli che caratterizzano l'attività lavorativa ed il rapporto tra questa e la formazione acquisita nel corso di dottorato. Tra i primi si segnalano l'*acquisizione di professionalità* (mediana: 8; asimmetria: -1,06; curtosi: 1,26), l'*indipendenza e autonomia sul lavoro* (mediana: 8; asimmetria: -1,09; curtosi: 1,14), i *rapporti coi colleghi* (mediana: 8; asimmetria: -1,12; curtosi: 1,61), la *flessibilità dell'orario di lavoro* (mediana: 8; asimmetria: -1,05;

<sup>125</sup> A ciascuna modalità è stato assegnato un punteggio al quale corrispondono i valori 1= per niente e 10 = moltissimo.

curtosi: 0,51) e la *localizzazione della sede di lavoro* (mediana: 8; asimmetria: -0,79; curtosi: 0,01). Tra gli altri si evidenziano, in particolare, la *coerenza con gli studi fatti* (mediana: 8; asimmetria: -1,02; curtosi: 0,63), l'*utilizzo delle competenze acquisite* (mediana: 8; asimmetria: -1,04; curtosi: 0,89) e la *rispondenza ai propri interessi culturali* (mediana: 8; asimmetria: -1,18; curtosi: 1,35). La distribuzione di frequenze di tutti questi aspetti è caratterizzata da un moderato grado di asimmetria negativa ed è indicativa di un addensamento di frequenze verso le modalità di risposta più elevate.

Gli aspetti che hanno ricevuto una valutazione peggiore sono quelli che descrivono le aspettative per il futuro, in termini di *prospettive di guadagno* (mediana: 5; asimmetria: -0,07; curtosi: -0,98), *possibilità di carriera* (mediana: 5; asimmetria: +0,01; -1,07) e *stabilità/sicurezza sul lavoro* (mediana: 5; asimmetria: +0,22; curtosi: -1,29), le cui distribuzioni risultano abbastanza simmetriche, e indicano un addensamento di frequenze verso le modalità di risposta centrali.

### C.3.10 Il livello di retribuzione attuale

Ad un anno dal conseguimento del titolo, il reddito mensile nominale netto<sup>126</sup> dichiarato è stato di 1.592,87 euro con uno scarto quadratico medio di 912,41 euro.

La maggior parte dei dottori di ricerca ha dichiarato (cfr. Tabella C.A32) un reddito mensile nominale netto inferiore a 1.500 euro (61,0%). In prevalenza, questi si collocano nella classe di reddito compresa tra 1.000 e 1.500 euro (48,6%) ma la percentuale dei redditi inferiori a 1.000 euro riguarda quasi un sesto del totale (12,1%) e non può considerarsi trascurabile mentre i redditi inferiori a 500 euro rappresentano una quota marginale (3,4%). I redditi mensili di almeno 2.000 euro sono poco meno di un quinto (18,4%) mentre quelli oltre i 3.000 euro sono soltanto uno su venti (4,8%). Nel complesso, la distribuzione del reddito presenta un certo grado di asimmetria positiva (mediana: classe € 1.001 - € 1.500; asimmetria: +1,59; curtosi: +3,87) ed è indicativa di una concentrazione di frequenze nelle modalità di risposta medio basse e della contemporanea presenza di poche frequenze nelle classi di reddito più distanti da queste.

---

<sup>126</sup> Il reddito è stato misurato chiedendo a ciascun intervistato di indicare il valore percepito scegliendo tra dieci classi: meno di € 500; € 500 - € 1.000; € 1.001 - € 1.500; € 1.501 - € 2.000; € 2.001 - € 2.500; € 2.501 - € 3.000; € 3.001 - € 3.500; € 3.501 - € 4.000; € 4.001 - € 6.000; oltre € 6.000. La media aritmetica è stata calcolata in maniera approssimata considerando al posto della classe il rispettivo valore centrale. Alla classe oltre € 6.000 è stato attribuito il valore 8.000 euro.

Differenze retributive significative si osservano per genere, area disciplinare e profilo giuridico del lavoro.

Le differenze retributive di genere sono a vantaggio dei *maschi* (1.731,30 euro) rispetto alle *femmine* (1.465,01 euro).

Per quanto riguarda le macro aree disciplinari, i dottori che guadagnano più della media sono quelli delle *scienze economiche (aziendali)* (2.179,25 euro), delle *scienze giuridiche* (2072,49 euro) e delle *scienze mediche* (1.953,68 euro). Livelli retributivi inferiori alla media si osservano invece tra i dottori delle *scienze storico-artistiche* (1.213,64 euro), delle *scienze dell'antichità e filologico-letterarie* (1.266,01 euro) e delle *scienze storiche e filosofiche* (1.278,69 euro).

Infine, per quanto riguarda il profilo giuridico, i *lavoratori autonomi* (1.720,67 euro) hanno migliori retribuzioni dei *lavoratori dipendenti* (1.576,41 euro). Tuttavia, tra questi ultimi, la variabilità nelle retribuzioni è piuttosto elevata. Livelli retributivi superiori alla media si registrano tra i *dirigenti* (2.782,37 euro) ed i *direttivi/quadri* (2.172,68 euro) mentre livelli inferiori alla media si osservano tra i *docenti delle scuole medie inferiori* (1250,00 euro), i *docenti delle scuole medie superiori* (1.270,16 euro) e gli *assegnisti/borsisti* (1.365,84 euro).

Il grado di soddisfazione rispetto al *livello di retribuzione attuale* è stato giudicato (cfr. Tabella C.A30.1) appena sufficiente (mediana: 6; asimmetria: -0,39; curtosi: -0,59). I più soddisfatti sono i dottori delle *scienze matematiche* (mediana: 7; asimmetria: -0,83; curtosi: -0,39), quelli meno soddisfatti sono i dottori *ingegneria civile e architettura* (mediana: 5; asimmetria: -0,19; curtosi: -0,72), *scienze psicologiche, geografiche e demoeoantropologiche* (mediana: 5; asimmetria: -0,13; curtosi: -1,01) e delle *scienze storico-artistiche* (mediana: 5; asimmetria: -0,20; curtosi: -0,69).

Nonostante il livello di soddisfazione rispetto alla retribuzione sia piuttosto limitato, coloro che *svolgono un'altra attività retribuita oltre all'attività principale* sono poco più di un quinto (21,7%) (cfr. Tabella C.A31), in particolare si tratta di *lavoratori autonomi* (35,0%) più che di *lavoratori dipendenti* (19,9%) mentre non si osservano differenze di genere significative (V di Cramer: 0,001; gdl: 1; p = 0,939).

### C.3.11 La ricerca di un nuovo lavoro

Negli ultimi tre mesi hanno cercato attivamente lavoro (cfr. Tabella C.A33) poco meno di un terzo dei dottori di ricerca (29,1%). Questa percentuale è risultata maggiore per i dottori in *scienze informatiche* (41,2%), *scienze della terra* (39,0%) e *scienze economiche (socio-politiche)* (38,2%) ed inferiore per i dottori in *scienze statistiche* (5,3%), *scienze pedagogiche* (12,5%), *scienze giuridiche* (15,5%) ed *ingegneria industriale* (22,4%) (cfr. Tabella C.A33a e Tabella C.A33b).

Nel complesso, il grado di associazione tra la ricerca di un nuovo lavoro e l'area disciplinare del corso di dottorato è piuttosto moderata (V di Cramer: 0,139; gdl: 22;  $p < 0,001$ ) mentre non si osservano differenze di genere significative (V di Cramer: 0,005; gdl: 1;  $p < 0,791$ ).

In generale, il livello di *soddisfazione nel complesso* per il lavoro svolto è risultato più basso tra chi ha cercato attivamente un lavoro negli ultimi tre mesi (mediana: 6; asimmetria: -0,50; curtosi: -0,14) rispetto a chi non ha cercato lavoro (mediana: 7; asimmetria: -0,71; curtosi: +1,24). Tuttavia, se si considera il livello di *soddisfazione rispetto alle mansioni svolte*, la differenza tra chi ha cercato lavoro (mediana: 6; asimmetria: -0,57; curtosi: -0,32) e chi non ha cercato lavoro (mediana: 8; asimmetria: -0,97; curtosi: +1,35) è ancora più evidente.

Tra gli aspetti specifici, la differenza più accentuata tra chi ha cercato lavoro e chi non ha cercato lavoro negli ultimi tre mesi si è registrata invece per la *stabilità/sicurezza sul lavoro* (mediana: 3; asimmetria: +0,76; curtosi: -0,67 contro mediana: 5; asimmetria: -0,02; curtosi: -1,30), le *possibilità di carriera* (mediana: 4; asimmetria: +0,43; curtosi: -0,90 contro mediana: 6; asimmetria: -0,16; curtosi: -0,96) e le *prospettive di guadagno* (mediana: 4; asimmetria: +0,36; curtosi: -0,87 contro mediana: 6; asimmetria: -0,22; curtosi: -0,83).

Il motivo principale che ha indotto ad avviare la ricerca di un nuovo lavoro è stato la *stabilità del posto di lavoro* (44,7%). Tale motivo è risultato più diffuso tra i dottori delle *scienze veterinarie* (71,4%), *scienze matematiche* (60,0%), *scienze fisiche* (58,3%), *scienze agrarie* (56,4%) e *scienze della terra* (56,1%) e tra le *femmine* (56,9%) rispetto ai *maschi* (43,1%).

Per quanto riguarda le altre motivazioni, una percentuale significativa è stata raggiunta solo dall'*insoddisfazione per il livello di retribuzione* (13,2%) e dalla *più*

*elevata possibilità di carriera (12,1%) mentre il rimanente è distribuito tra la possibilità di acquisire maggiore professionalità (6,9%), la maggiore coerenza con gli studi fatti (6,0%), la localizzazione della sede di lavoro (5,2%), la maggiore rispondenza al titolo posseduto (5,1%), la maggiore rispondenza ai propri interessi culturali (4,1%) ed altre motivazioni ancora (2,6%).*

#### **C.4 Chi attualmente non lavora ma ha lavorato dopo il conseguimento del titolo**

##### *C.4.1 L'ultimo lavoro e gli aspetti migliorati dopo il conseguimento del titolo*

L'ultimo lavoro svolto da chi non lavora attualmente ma ha lavorato dopo il conseguimento del titolo (cfr. Tabella C.B01) è stato, in prevalenza, lo stesso che svolgevano al momento del conseguimento del titolo (54,5%). Questa percentuale è risultata maggiore soprattutto per i dottori nelle *scienze giuridiche* (94,1%).

*Il conseguimento del titolo di dottore di ricerca ha comportato un miglioramento nel proprio lavoro* per poco più di un quarto dei dottori (28,7%) (cfr. Tabella C.B04).

Gli aspetti che hanno fatto registrare un miglioramento significativo dopo il conseguimento del titolo di dottore sono stati *l'utilizzo delle competenze acquisite* (92,9%), *la rispondenza ai propri interessi culturali* (92,9%), e *l'acquisizione di professionalità* (86,2%) (cfr. Tabelle C.B04b-C.B04b16).

Non si registrano effetti positivi sul *tempo libero* (3,7%), sulla *stabilità/sicurezza sul lavoro* (11,1%) e sulla *localizzazione della sede di lavoro* (22,2%).

##### *C.4.2 L'interruzione dell'ultima attività lavorativa*

Il motivo principale dell'interruzione dell'ultima attività lavorativa è stato (cfr. Tabella C.B06) *la scadenza regolare di un contratto a tempo determinato o termine di un lavoro occasionale o stagionale* che ha riguardato quasi la metà dei dottori (49,9%).

Una quota significativa di dottori ha invece dichiarato di essersi dimesso (12,1%), in particolare *perché non soddisfatto di quel tipo di lavoro/contratto* (5,8%), *perché non soddisfatto del trattamento economico* (3,1%), *per motivi personali* (2,5%) o *perché non soddisfatto della sede di lavoro* (0,7%).

Per quanto riguarda le altre motivazioni, uno su dieci ha dichiarato di aver

interrotto l'attività lavorativa per *iniziare un'attività di formazione post-dottorato* (9,2%) o per *maternità* (5,6%).

Un quarto dei dottori ha dichiarato la modalità “altro motivo” (21,1%).

#### *C.4.3 La ricerca di un nuovo lavoro*

*Negli ultimi tre mesi hanno cercato attivamente un lavoro* (cfr. Tabella C.B07) poco più della metà dei dottori che hanno perso il posto (58,4%). I più attivi nella ricerca di un nuovo lavoro sono stati i dottori delle *scienze economiche (socio-politiche)* (88,9%), *scienze storiche e filosofiche* (67,6%), *scienze storico-artistiche* (66,7%) e *scienze agrarie* (65,2%) e, per quanto riguarda il genere, le *femmine* (60,1%) rispetto ai *maschi* (39,9%).

Il tipo di lavoro cercato è stato in prevalenza quello di *ricercatore o docente universitario* (31,6%) ma più di un quarto dei dottori ha dichiarato di *non avere preferenze* (26,4%) (cfr. Tabella C.B08).

Il rimanente dei dottori ha cercato un lavoro da *libero professionista (iscritto albo)* (11,1%), *docente alle scuole medie (inferiori o superiori)* (9,4%), *ricercatore nel settore pubblico non universitario* (6,5%), *lavoro autonomo* (5,2%), *impiegato o intermedio* (5,2%), *direttivo/quadro* (2,9%) o *dirigente* (1,6%).

Gli aspetti ritenuti più importanti nella ricerca del lavoro sono stati la *stabilità/sicurezza sul lavoro* (mediana: 9; asimmetria: -1,54; curtosi: 3,75), l'*utilizzo delle competenze acquisite* (mediana: 9; asimmetria: -0,61; curtosi: -0,51) e l'*acquisizione di professionalità* (mediana: 9; asimmetria: -1,01; curtosi: +0,36) (Tabella C.B30c).

### **C.5 Chi non ha mai lavorato dopo il conseguimento del titolo**

#### *C.5.1 La ricerca del primo lavoro*

La maggior parte dei dottori che non hanno mai lavorato dopo il conseguimento del titolo di dottore di ricerca ha dichiarato (cfr. Tabella C.C01) di *essere attualmente alla ricerca di un lavoro* (81,1%).

Il tipo di lavoro cercato è stato in prevalenza quello di *ricercatore o docente*

*universitario* (50,5%) ma quasi un terzo dei dottori ha dichiarato di *non avere preferenze* (29,5%).

Gli aspetti ritenuti più importanti nella ricerca del lavoro sono stati la *stabilità/sicurezza sul lavoro* (mediana: 9; asimmetria: -0,72; curtosi: -0,27), l'*utilizzo delle competenze acquisite* (mediana: 9; asimmetria: -0,71; curtosi: -0,25), l'*acquisizione di professionalità* (mediana: 9; asimmetria: -1,52; curtosi: +3,60), la *coerenza con gli studi fatti* (mediana: 9; asimmetria: -0,55; curtosi: -0,92), la *rispondenza ai propri interessi culturali* (mediana: 9; asimmetria: -0,66; curtosi: -0,62) ed il *riconoscimento del titolo di dottore di ricerca* (mediana: 9; asimmetria: -1,29; curtosi: +1,48) (Tabella C.C03c).

#### C.5.2 I motivi della non ricerca

La maggior parte dei dottori che non hanno cercato lavoro lo hanno fatto (cfr. Tabella C.04) per *motivi di studio o di formazione post-dottorato* (56,5%) o per *motivi personali* (*casalinga/o, maternità, assistenza figli o parenti, problemi di salute, pensione, ecc.*) (21,7%).

Il rimanente non hanno cercato lavoro perché in *preparazione ad un concorso pubblico* (8,7%), perché *in attesa di avviare un'attività in proprio* (8,7%) o perché *in attesa di chiamata dopo aver superato prova / concorso / selezione* (4,3%).



## Blocco Z: tutti i dottori di ricerca

**Tabella C.Z01**

*Qual è il motivo principale per cui ha deciso di iscriversi al Corso di Dottorato in cui ha conseguito il titolo di Dottore di Ricerca?*

	N	%	% cum
Per completare/arricchire la mia formazione	817	19,5	19,5
Per conseguire un titolo dal più alto potere professionalizzante	436	10,4	29,9
Interesse per lo studio	268	6,4	36,3
Per fare attività di ricerca	1735	41,4	77,8
Per l'interesse nei confronti di una futura carriera accademica	849	20,3	98,0
Altro	82	2,0	100,0
<b>Totale validi</b>	<b>4187</b>	<b>100,0</b>	
Mancanti	36		
Totale assoluto	4223		

**Tabella C.Z02**

*Durante il Dottorato in cui ha conseguito il titolo di Dottore di Ricerca Lei è stato titolare di:*

	N	%	% cum
Un posto con borsa finanziata dall'Università/MIUR	2442	57,9	57,9
Un posto con borsa finanziata da enti pubblici	275	6,5	64,4
Un posto con borsa finanziata da enti privati	220	5,2	69,6
Un posto senza borsa ma coperto da assegno di ricerca	258	6,1	75,7
Un posto senza borsa e non coperto da assegno di ricerca	758	18,0	93,7
Altro	268	6,3	100,0
<b>Totale validi</b>	<b>4221</b>	<b>100,0</b>	
Mancanti	2		
Totale assoluto	4223		

*Di seguito Le vengono elencati una serie di aspetti della formazione ricevuta durante il Corso di Dottorato.*

*Per ciascuno mi dovrebbe dire quanto è soddisfatto, su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo).*

**Tabella C.Z03.1**

*Contenuto degli insegnamenti*

	N	%	% cum
1 = Per niente	352	8,4	8,4
2	211	5,0	13,4
3	267	6,3	19,7
4	303	7,2	26,9
5	452	10,7	37,7
6	651	15,5	53,1
7	723	17,2	70,3
8	726	17,3	87,6
9	302	7,2	94,8
10 = Moltissimo	220	5,2	100,0
<b>Totale validi</b>	<b>4207</b>	<b>100,0</b>	
Mancanti	16		
Totale assoluto	4223		

Tabella C.Z03.2

Qualità della didattica del personale docente

	N	%	% cum
1 = Per niente	300	7,1	7,1
2	170	4,0	11,2
3	206	4,9	16,1
4	213	5,1	21,1
5	378	9,0	30,1
6	615	14,6	44,7
7	735	17,5	62,2
8	875	20,8	83,0
9	435	10,3	93,4
10 = Moltissimo	279	6,6	100,0
<b>Totale validi</b>	<b>4206</b>	<b>100,0</b>	
Mancanti	17		
Totale assoluto	4223		

Tabella C.Z03.3

Qualità scientifica di professori/ricercatori

	N	%	% cum
1 = Per niente	105	2,5	2,5
2	74	1,8	4,3
3	113	2,7	6,9
4	129	3,1	10,0
5	238	5,7	15,6
6	453	10,8	26,4
7	762	18,1	44,5
8	1060	25,2	69,7
9	775	18,4	88,1
10 = Moltissimo	502	11,9	100,0
<b>Totale validi</b>	<b>4211</b>	<b>100,0</b>	
Mancanti	12		
Totale assoluto	4223		

Tabella C.Z03.4

Sviluppo / addestramento della capacità di fare ricerca

	N	%	% cum
1 = Per niente	238	5,7	5,7
2	154	3,7	9,3
3	187	4,4	13,8
4	234	5,6	19,3
5	389	9,2	28,6
6	557	13,2	41,8
7	670	15,9	57,7
8	773	18,4	76,1
9	563	13,4	89,5
10 = Moltissimo	442	10,5	100,0
<b>Totale validi</b>	<b>4207</b>	<b>100,0</b>	
Mancanti	16		
Totale assoluto	4223		

Tabella C.Z03.5

Possibilità di pubblicare (articoli, volumi, ...)

	N	%	% cum
1 = Per niente	397	9,4	9,4
2	214	5,1	14,5
3	225	5,3	19,9
4	266	6,3	26,2
5	466	11,1	37,2
6	528	12,5	49,8
7	621	14,8	64,5
8	645	15,3	79,9
9	442	10,5	90,4
10 = Moltissimo	406	9,6	100,0
<b>Totale validi</b>	<b>4210</b>	<b>100</b>	
Mancanti	13		
Totale assoluto	4223		

Tabella C.Z04

Se potesse tornare indietro, si iscriverebbe di nuovo ad un Corso di Dottorato?

	N	%	% cum
Sì, allo stesso Corso nella stessa sede universitaria	2170	52,9	52,9
Sì, allo stesso Corso ma in altra sede universitaria	772	18,8	71,7
Sì ma ad altro Corso	479	11,7	83,3
No	684	16,7	100,0
<b>Totale validi</b>	<b>4105</b>	<b>100,0</b>	
Mancanti	118		
Totale assoluto	4223		

Tabella C.Z05

Durante il Corso di Dottorato ha compiuto periodi di studio all'estero?

	N	%	% cum
Sì	1644	39,6	39,6
No	2507	60,4	100,0
<b>Totale validi</b>	<b>4151</b>	<b>100,0</b>	
Mancanti	72		
Totale assoluto	4223		

Tabella C.Z06

Durante il Corso di Dottorato ha svolto una qualche attività lavorativa (inclusi co.co.co. e co.co.pro.)?

	N	%	% cum
Sì, attività attinente all'attività di ricerca del dottorato	1435	34,3	34,3
Sì, attività non attinente all'attività di ricerca del dotto	1007	24,1	58,4
No	1738	41,6	100,0
<b>Totale validi</b>	<b>4180</b>	<b>100,0</b>	
Mancanti	43		
Totale assoluto	4223		

Tabella C.Z07

Lei lavorava al momento del conseguimento del titolo di Dottore di Ricerca?

	N	%	% cum
Sì	1968	46,7	46,7
No	2249	53,3	100,0
<b>Totale validi</b>	<b>4217</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	4223		

Tabella C.Z08

Indichi la sua condizione occupazionale attuale: oggi

	N	%	% cum
Lavoro con assegno di ricerca, borsa post-doc o altra borsa	1476	35,4	35,4
Lavoro con altra forma contrattuale a tempo determinato (con	896	21,5	56,8
Lavoro a tempo indeterminato / autonomo effettivo	1124	26,9	83,8
Attività di formazione post-dottorato retribuita	77	1,8	85,6
Attività di formazione post-dottorato non retribuita	69	1,7	87,3
Periodo di studio in preparazione a concorsi pubblici/esame	47	1,1	88,4
Ricerca di un lavoro	239	5,7	94,1
Altra condizione di inattività (non lavora e non cerca lavoro)	19	0,5	94,6
Altro	226	5,4	100,0
<b>Totale validi</b>	<b>4173</b>	<b>100</b>	
Mancanti	50		
<b>Totale assoluto</b>	<b>4223</b>		

## Blocco A: chi attualmente lavora

**Tabella C.A02**

*Il lavoro attuale è lo stesso che svolgeva al momento del conseguimento del titolo di Dottore? (solo se Z07 = 1)*

	N	%	% cum
Si	1195	69,3	69,3
No	529	30,7	100,0
<b>Totale validi</b>	<b>1724</b>	<b>100,0</b>	
Mancanti	1764		
Totale assoluto	3488		

**Tabella C.A03 (solo se Z07 = 1)**

*Il conseguimento del titolo di Dottore di Ricerca ha comportato un miglioramento nel suo lavoro? (Z07=1 e A02=1)*

	N	%	% cum
Si	450	37,8	37,8
No	742	62,2	100,0
<b>Totale validi</b>	<b>1192</b>	<b>100,0</b>	
Mancanti	3		
Totale assoluto	1195		

*Quali aspetti del suo lavoro sono migliorati? (Z07=1 e A02=1)*

**Tabella C.A06.1**

*Livello di retribuzione attuale*

	N	%	% cum
Si	206	51,8	51,8
No	192	48,2	100,0
<b>Totale validi</b>	<b>398</b>	<b>100,0</b>	
Mancanti	52		
Totale assoluto	450		

**Tabella C.A06.2**

*Prospettive di guadagno*

	N	%	% cum
Si	180	46,4	46,4
No	208	53,6	100,0
<b>Totale validi</b>	<b>388</b>	<b>100,0</b>	
Mancanti	62		
Totale assoluto	450		

**Tabella C.A06.3**

*Possibilità di carriera*

	N	%	% cum
Si	292	70,7	70,7
No	121	29,3	100,0
<b>Totale validi</b>	<b>413</b>	<b>100,0</b>	
Mancanti	37		
Totale assoluto	450		

Tabella C.A06.4

Stabilità/sicurezza sul lavoro

	N	%	% cum
Si	83	21,1	21,1
No	311	78,9	100,0
<b>Totale validi</b>	<b>394</b>	<b>100,0</b>	
Mancanti	56		
Totale assoluto	450		

Tabella C.A06.5

Coerenza con studi fatti

	N	%	% cum
Si	335	80,9	80,9
No	79	19,1	100,0
<b>Totale validi</b>	<b>414</b>	<b>100,0</b>	
Mancanti	36		
Totale assoluto	450		

Tabella C.A06.6

Utilizzo delle competenze acquisite

	N	%	% cum
Si	388	90,4	90,4
No	41	9,6	100,0
<b>Totale validi</b>	<b>429</b>	<b>100,0</b>	
Mancanti	21		
Totale assoluto	450		

Tabella C.A06.7

Acquisizione di professionalità

	N	%	% cum
Si	383	91,2	91,2
No	37	8,8	100,0
<b>Totale validi</b>	<b>420</b>	<b>100,0</b>	
Mancanti	30		
Totale assoluto	450		

Tabella C.A06.8

Rispondenza ai propri interessi culturali

	N	%	% cum
Si	345	85,0	85,0
No	61	15,0	100,0
<b>Totale validi</b>	<b>406</b>	<b>100,0</b>	
Mancanti	44		
Totale assoluto	450		

Tabella C.A06.9

Indipendenza e autonomia sul lavoro

	N	%	% cum
Si	261	66,2	66,2
No	133	33,8	100,0
<b>Totale validi</b>	<b>394</b>	<b>100,0</b>	
Mancanti	56		
Totale assoluto	450		

Tabella C.A06.10

*Coinvolgimento nei processi decisionali*

	N	%	% cum
Si	219	56,2	56,2
No	171	43,8	100,0
<b>Totale validi</b>	<b>390</b>	<b>100,0</b>	
Mancanti	60		
Totale assoluto	450		

Tabella C.A06.11

*Flessibilità dell'orario di lavoro*

	N	%	% cum
Si	137	35,1	35,1
No	253	64,9	100,0
<b>Totale validi</b>	<b>390</b>	<b>100,0</b>	
Mancanti	60		
Totale assoluto	450		

Tabella C.A06.12

*Tempo libero*

	N	%	% cum
Si	47	12,2	12,2
No	339	87,8	100,0
<b>Totale validi</b>	<b>386</b>	<b>100,0</b>	
Mancanti	64		
Totale assoluto	450		

Tabella C.A06.13

*Localizzazione sede di lavoro*

	N	%	% cum
Si	72	18,7	18,7
No	314	81,3	100,0
<b>Totale validi</b>	<b>386</b>	<b>100,0</b>	
Mancanti	64		
Totale assoluto	450		

Tabella C.A06.14

*Rapporti coi colleghi*

	N	%	% cum
Si	155	39,3	39,3
No	239	60,7	100,0
<b>Totale validi</b>	<b>394</b>	<b>100,0</b>	
Mancanti	56		
Totale assoluto	450		

Tabella C.A06.15

*Miglioramento generale rispetto alle aspettative iniziali*

	N	%	% cum
Si	226	57,5	57,5
No	167	42,5	100,0
<b>Totale validi</b>	<b>393</b>	<b>100,0</b>	
Mancanti	57		
Totale assoluto	450		

Tabella C.A06.16

Miglioramento generale rispetto all'attività lavorativa ideale

	N	%	% cum
Si	174	44,6	44,6
No	216	55,4	100,0
<b>Totale validi</b>	<b>390</b>	<b>100,0</b>	
Mancanti	60		
Totale assoluto	450		

Tabella C.A10

In base a quale canale ha trovato l'attuale lavoro?

	N	%	% cum
Lettura di offerte impiego sui giornali	14	0,6	0,6
Risposta a inviti a partecipare a colloqui da parte di aziende	33	1,5	2,2
Contatti con datori di lavoro su segnalazione università/docenti	368	17,1	19,2
Contatti con datori di lavoro su iniziativa personale o altra iniziativa	359	16,6	35,9
Domande o partecipazione a concorsi pubblici	711	33,0	68,8
Proseguendo un'attività familiare già esistente	15	0,7	69,5
Domande a provveditorati o presidi di insegnamento	61	2,8	72,4
Tramite siti internet e/o motori di ricerca	161	7,5	79,8
Altro	435	20,2	100,0
<b>Totale validi</b>	<b>2157</b>	<b>100,0</b>	
Mancanti	1331		
Totale assoluto	3488		

Tabella C.A11

Il suo attuale lavoro è di tipo autonomo o dipendente?

	N	%	% cum
Autonomo	444	12,8	12,8
Dipendente inclusi contratti di collaborazione	3017	87,2	100,0
<b>Totale validi</b>	<b>3461</b>	<b>100,0</b>	
Mancanti	27		
Totale assoluto	3488		

Tabella C.A12

Con che tipo di contratto è regolato il lavoro da Lei attualmente svolto? (A11 = 2)

	N	%	% cum
Tempo pieno e indeterminato	840	27,9	27,9
Part-time, tempo indeterminato	51	1,7	29,6
Tempo pieno, a termine	441	14,6	44,2
Part-time, a termine	68	2,3	46,5
Collaborazione coordinata e continuativa	140	4,6	51,1
Collaborazione a progetto	226	7,5	58,7
Collaborazione occasionale	34	1,1	59,8
Prestazione d'opera	24	0,8	60,6
Assegno di ricerca o borsa di studio	1124	37,3	97,9
Non è regolato da contratto	19	0,6	98,5
Altro	44	1,5	100,0
<b>Totale validi</b>	<b>3011</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	3017		



Tabella C.A14

Può qualificarmi sotto il profilo giuridico il suo lavoro attuale? (A11 = 2)

	N	%	% cum
Dirigente	149	5,0	5,0
Direttivo/quadro	105	3,5	8,5
Impiegato o intermedio	466	15,6	24,1
Ricercatore nel settore privato	133	4,5	28,6
Assegnista/borsista	1180	39,5	68,1
Ricercatore universitario a tempo determinato	161	5,4	73,5
Ricercatore universitario di ruolo	104	3,5	77,0
Professore universitario associato	8	0,3	77,3
Professore universitario ordinario	4	0,1	77,4
Ricercatore (settore pubblico non universitario)	164	5,5	82,9
Primo ricercatore (settore pubblico non universitario)	5	0,2	83,0
Dirigente di ricerca (settore pubblico non universitario)	4	0,1	83,2
Docente alle scuole medie inferiori	67	2,2	85,4
Docente alle scuole medie superiori	136	4,6	90,0
Altro	299	10,0	100,0
<b>Totale validi</b>	<b>2985</b>	<b>100,0</b>	
Mancanti	32		
Totale assoluto	3017		

Tabella C.A15

Può qualificarmi sotto il profilo giuridico il suo lavoro attuale? (A11=1)

	N	%	% cum
Imprenditore	26	5,9	5,9
Libero professionista iscritto albo	290	66,4	72,3
Libero professionista non iscritto albo	121	27,7	100,0
<b>Totale validi</b>	<b>437</b>	<b>100,0</b>	
Mancanti	7		
Totale assoluto	444		

Tabella C.A16

Nel suo lavoro svolge mansioni di ricerca e sviluppo? (A14<=3 o A11=1)

	N	%	% cum
Sì, come attività principale	100	13,9	13,9
Sì, ma non come attività principale	306	42,4	56,3
No	315	43,7	100,0
<b>Totale validi</b>	<b>721</b>	<b>100,0</b>	
Mancanti	443		
Totale assoluto	1164		

Tabella C.A17

Nel suo lavoro svolge attività di innovazione? (A14<=3 o A11=1)

	N	%	% cum
Sì, sia innovazione di processo che innovazione di prodotto	137	19,4	19,4
Sì, innovazione di processo	121	17,1	36,4
Sì, innovazione di prodotto	65	9,2	45,6
No	385	54,4	100,0
<b>Totale validi</b>	<b>708</b>	<b>100,0</b>	
Mancanti	456		
Totale assoluto	1164		

**Tabella C.A18**

*Nel suo lavoro svolge attività di innovazione? (A14>=5 e A14<=9)*

	N	%	% cum
Nello stesso Ateneo in cui ha svolto il dottorato	874	64,4	64,4
In un Ateneo diverso ma nella stessa regione	84	6,2	70,5
In un Ateneo diverso di altra regione italiana	200	14,7	85,3
All'estero	200	14,7	100,0
<b>Totale validi</b>	<b>1358</b>	<b>100,0</b>	
Mancanti	100		
Totale assoluto	1458		

**Tabella C.A19**

*Le attività di ricerca svolte nella posizione attuale sono coerenti con la formazione acquisita nel Corso di Dottorato in cui ha conseguito il titolo? ((A14>=4 e A14<=12) o A16<=2)*

	N	%	% cum
Pienamente coerente	1150	67,3	67,3
Parzialmente coerente	482	28,2	95,5
Non coerente	77	4,5	100,0
<b>Totale validi</b>	<b>1709</b>	<b>100,0</b>	
Mancanti	461		
Totale assoluto	2170		

**Tabella C.A24**

*Lavora in: (A14<4)*

	N	%	% cum
Ente pubblico	316	44,3	44,3
Azienda a partecipazione pubblica	64	9,0	53,2
Azienda privata italiana (anche se in filiale all'estero)	228	31,9	85,2
Azienda privata straniera (anche se in filiale in Italia)	82	11,5	96,6
Organizzazione no-profit	24	3,4	100,0
<b>Totale validi</b>	<b>714</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	720		

**Tabella C.A25**

*Quanti dipendenti ha l'azienda per cui lavora? (A24>1)*

	N	%	% cum
Fino a 10 dipendenti	67	17,1	17,1
Da 11 a 50 dipendenti	68	17,3	34,4
Da 51 a 250 dipendenti	79	20,2	54,6
Più di 250 dipendenti	178	45,4	100,0
<b>Totale validi</b>	<b>392</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	398		

**Tabella C.A25**

*Può indicarmi il ramo di attività economica dell'azienda per cui lavora: (A14<5)*

	N	%	% cum
Agricoltura - foreste - caccia - pesca	18	2,2	2,2
Stampa ed editoria	10	1,2	3,4
Energia, Gas, Acqua	34	4,1	7,5
Estrazione mineraria	0	0,0	7,5
Chimica	57	6,9	14,4
Metalmeccanica e Mecc. precisione	32	3,9	18,3
Alimentari e Tabacchi	9	1,1	19,4
Tessile, Abbigl, Cuoio, Calzature	5	0,6	20,0
Legno, Arredamento	5	0,6	20,6
Carta, Gomma, Plastica	1	0,1	20,7
Costruz.- installa. d'impianti	19	2,3	23,0
Elettronica e Elettrotecnica	27	3,3	26,3
Manifattura varia	1	0,1	26,4
Commercio, Alberghi, Esercizi Pubbl.	9	1,1	27,5
Trasporti	18	2,2	29,7
Comunicazioni e Telecomunicazioni	13	1,6	31,2
Credito e Assicurazioni	34	4,1	35,4
Consulenza legale, amm.va, contabile	12	1,5	36,8
Pubblicità, pubbliche relazioni	1	0,1	36,9
Informatica, elaborazione dati	39	4,7	41,6
Servizi alle imprese	30	3,6	45,3
Servizi ricreativi e culturali	13	1,6	46,9
Pubblica Amministrazione, Forze Armate	106	12,8	59,7
Istruzione, ricerca e sviluppo	107	13,0	72,6
Sanità	195	23,6	96,2
Altri Servizi Sociali	31	3,8	100,0
<b>Totale validi</b>	<b>826</b>	<b>100,0</b>	
Mancanti	27		
<b>Totale assoluto</b>	<b>853</b>		

**Tabella C.A25**

*Quanto le è utile sul lavoro la formazione ricevuta durante il corso di dottorato?*

	N	%	% cum
È fondamentale per le mansioni che svolgo	1291	39,6	39,6
È utile per il mio approccio al lavoro, anche se non ne faccio	1264	38,7	78,3
Ne faccio un uso piuttosto limitato	406	12,4	90,7
Non ne faccio alcun uso	303	9,3	100,0
<b>Totale validi</b>	<b>3264</b>	<b>100,0</b>	
Mancanti	224		
<b>Totale assoluto</b>	<b>3488</b>		

Può esprimere su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo) il suo livello di soddisfazione per i seguenti aspetti del lavoro attuale?

**Tabella C.A30.1**

*Livello di retribuzione attuale*

	N	%	% cum
1 = Per niente	268	8,0	8,0
2	164	4,9	12,9
3	262	7,8	20,7
4	274	8,2	28,9
5	438	13,1	42,0
6	595	17,8	59,7
7	629	18,8	78,5
8	462	13,8	92,3
9	159	4,7	97,0
10 = Moltissimo	100	3,0	100,0
<b>Totale validi</b>	<b>3351</b>	<b>100,0</b>	
Mancanti	137		
<b>Totale assoluto</b>	<b>3488</b>		

**Tabella C.A30.2**

*Prospettive di guadagno*

	N	%	% cum
1 = Per niente	456	13,6	13,6
2	239	7,1	20,8
3	314	9,4	30,1
4	325	9,7	39,9
5	453	13,5	53,4
6	510	15,2	68,6
7	403	12,0	80,7
8	389	11,6	92,3
9	153	4,6	96,9
10 = Moltissimo	105	3,1	100,0
<b>Totale validi</b>	<b>3347</b>	<b>100,0</b>	
Mancanti	141		
<b>Totale assoluto</b>	<b>3488</b>		

**Tabella C.A30.3**

*Possibilità di carriera*

	N	%	% cum
1 = Per niente	552	16,5	16,5
2	282	8,4	24,9
3	307	9,2	34,1
4	303	9,1	43,2
5	402	12,0	55,2
6	498	14,9	70,1
7	413	12,4	82,5
8	347	10,4	92,9
9	133	4,0	96,8
10 = Moltissimo	106	3,2	100,0
<b>Totale validi</b>	<b>3343</b>	<b>100,0</b>	
Mancanti	145		
<b>Totale assoluto</b>	<b>3488</b>		

Tabella C.A30.4

Stabilità/sicurezza sul lavoro

	N	%	% cum
1 = Per niente	830	24,8	24,8
2	279	8,3	33,1
3	270	8,1	41,2
4	258	7,7	48,9
5	316	9,4	58,4
6	314	9,4	67,7
7	285	8,5	76,2
8	306	9,1	85,4
9	244	7,3	92,7
10 = Moltissimo	245	7,3	100,0
<b>Totale validi</b>	<b>3347</b>	<b>100,0</b>	
Mancanti	141		
Totale assoluto	3488		

Tabella C.A30.5

Coerenza con studi fatti

	N	%	% cum
1 = Per niente	93	2,8	2,8
2	65	1,9	4,7
3	102	3,0	7,8
4	109	3,3	11,0
5	192	5,7	16,8
6	345	10,3	27,1
7	498	14,9	41,9
8	748	22,3	64,3
9	521	15,6	79,8
10 = Moltissimo	676	20,2	100,0
<b>Totale validi</b>	<b>3349</b>	<b>100,0</b>	
Mancanti	139		
Totale assoluto	3488		

Tabella C.A30.6

Utilizzo delle competenze acquisite

	N	%	% cum
1 = Per niente	82	2,4	2,4
2	55	1,6	4,1
3	95	2,8	6,9
4	86	2,6	9,5
5	210	6,3	15,8
6	345	10,3	26,1
7	581	17,3	43,4
8	787	23,5	66,9
9	526	15,7	82,6
10 = Moltissimo	582	17,4	100,0
<b>Totale validi</b>	<b>3349</b>	<b>100,0</b>	
Mancanti	139		
Totale assoluto	3488		

Tabella C.A30.7

Acquisizione di professionalità

	N	%	% cum
1 = Per niente	70	2,1	2,1
2	40	1,2	3,3
3	81	2,4	5,7
4	82	2,5	8,2
5	207	6,2	14,4
6	365	10,9	25,3
7	644	19,3	44,5
8	872	26,1	70,6
9	528	15,8	86,4
10 = Moltissimo	455	13,6	100,0
<b>Totale validi</b>	<b>3344</b>	<b>100,0</b>	
Mancanti	144		
Totale assoluto	3488		

Tabella C.A30.8

Rispondenza ai propri interessi culturali

	N	%	% cum
1 = Per niente	78	2,3	2,3
2	40	1,2	3,5
3	68	2,0	5,6
4	98	2,9	8,5
5	168	5,0	13,5
6	282	8,4	21,9
7	542	16,2	38,1
8	816	24,4	62,5
9	598	17,9	80,4
10 = Moltissimo	656	19,6	100,0
<b>Totale validi</b>	<b>3346</b>	<b>100,0</b>	
Mancanti	142		
Totale assoluto	3488		

Tabella C.A30.9

Indipendenza e autonomia sul lavoro

	N	%	% cum
1 = Per niente	80	2,4	2,4
2	53	1,6	4,0
3	75	2,2	6,2
4	90	2,7	8,9
5	193	5,8	14,7
6	364	10,9	25,6
7	578	17,3	42,9
8	818	24,5	67,3
9	583	17,4	84,7
10 = Moltissimo	510	15,3	100,0
<b>Totale validi</b>	<b>3344</b>	<b>100,0</b>	
Mancanti	144		
Totale assoluto	3488		

Tabella C.A30.10

*Coinvolgimento nei processi decisionali*

	N	%	% cum
1 = Per niente	203	6,1	6,1
2	129	3,9	9,9
3	132	3,9	13,9
4	175	5,2	19,1
5	298	8,9	28,0
6	460	13,8	41,8
7	612	18,3	60,1
8	671	20,1	80,2
9	378	11,3	91,5
10 = Moltissimo	285	8,5	100,0
<b>Totale validi</b>	<b>3343</b>	<b>100,0</b>	
Mancanti	145		
Totale assoluto	3488		

Tabella C.A30.11

*Flessibilità dell'orario di lavoro*

	N	%	% cum
1 = Per niente	141	4,2	4,2
2	76	2,3	6,5
3	95	2,8	9,3
4	116	3,5	12,8
5	195	5,8	18,6
6	324	9,7	28,3
7	462	13,8	42,1
8	731	21,9	64,0
9	562	16,8	80,8
10 = Moltissimo	643	19,2	100,0
<b>Totale validi</b>	<b>3345</b>	<b>100,0</b>	
Mancanti	143		
Totale assoluto	3488		

Tabella C.A30.12

*Tempo libero*

	N	%	% cum
1 = Per niente	167	5,0	5,0
2	216	6,5	11,4
3	238	7,1	18,6
4	288	8,6	27,2
5	424	12,7	39,8
6	562	16,8	56,6
7	517	15,4	72,1
8	485	14,5	86,6
9	229	6,8	93,4
10 = Moltissimo	221	6,6	100,0
<b>Totale validi</b>	<b>3347</b>	<b>100,0</b>	
Mancanti	141		
Totale assoluto	3488		

Tabella C.A30.13

Localizzazione sede di lavoro

	N	%	% cum
1 = Per niente	109	3,3	3,3
2	69	2,1	5,3
3	143	4,3	9,6
4	155	4,6	14,2
5	258	7,7	21,9
6	412	12,3	34,2
7	468	14,0	48,2
8	690	20,6	68,9
9	464	13,9	82,7
10 = Moltissimo	578	17,3	100,0
<b>Totale validi</b>	<b>3346</b>	<b>100,0</b>	
Mancanti	142		
Totale assoluto	3488		

Tabella C.A30.14

Rapporti coi colleghi

	N	%	% cum
1 = Per niente	41	1,2	1,2
2	32	1,0	2,2
3	43	1,3	3,5
4	83	2,5	6,0
5	168	5,0	11,0
6	345	10,3	21,3
7	582	17,4	38,8
8	856	25,6	64,4
9	691	20,7	85,1
10 = Moltissimo	498	14,9	100,0
<b>Totale validi</b>	<b>3339</b>	<b>100,0</b>	
Mancanti	149		
Totale assoluto	3488		

Tabella C.A30.15

Rispetto alle aspettative iniziali

	N	%	% cum
1 = Per niente	175	5,2	5,2
2	115	3,4	8,7
3	145	4,3	13,0
4	179	5,4	18,4
5	351	10,5	28,9
6	564	16,9	45,8
7	576	17,3	63,0
8	669	20,0	83,1
9	349	10,5	93,5
10 = Moltissimo	216	6,5	100,0
<b>Totale validi</b>	<b>3339</b>	<b>100,0</b>	
Mancanti	149		
Totale assoluto	3488		



Tabella C.A30.16

Rispetto alle mansioni svolte

	N	%	% cum
1 = Per niente	92	2,8	2,8
2	62	1,9	4,6
3	103	3,1	7,7
4	130	3,9	11,6
5	254	7,6	19,2
6	493	14,8	34,0
7	712	21,4	55,4
8	821	24,6	80,1
9	420	12,6	92,7
10 = Moltissimo	244	7,3	100,0
<b>Totale validi</b>	<b>3331</b>	<b>100,0</b>	
Mancanti	157		
Totale assoluto	3488		

Tabella C.A30.17

Rispetto al titolo di Dottore di Ricerca

	N	%	% cum
1 = Per niente	359	10,8	10,8
2	171	5,1	15,9
3	189	5,7	21,6
4	209	6,3	27,9
5	315	9,5	37,4
6	440	13,2	50,6
7	466	14,0	64,6
8	560	16,8	81,4
9	338	10,2	91,6
10 = Moltissimo	279	8,4	100,0
<b>Totale validi</b>	<b>3326</b>	<b>100,0</b>	
Mancanti	162		
Totale assoluto	3488		

Tabella C.A30.18

E nel suo complesso

	N	%	% cum
1 = Per niente	50	1,5	1,5
2	51	1,5	3,0
3	87	2,6	5,7
4	146	4,4	10,0
5	329	9,9	19,9
6	605	18,2	38,1
7	833	25,1	63,2
8	797	24,0	87,2
9	305	9,2	96,4
10 = Moltissimo	121	3,6	100,0
<b>Totale validi</b>	<b>3324</b>	<b>100,0</b>	
Mancanti	164		
Totale assoluto	3488		

**Tabella C.A31**

*Oltre all'attività principale Lei svolge un'altra attività retribuita?*

	N	%	% cum
Sì	712	21,7	21,7
No	2566	78,3	100,0
<b>Totale validi</b>	<b>3278</b>	<b>100,0</b>	
Mancanti	210		
Totale assoluto	3488		

**Tabella C.A32**

*Può indicare il reddito mensile netto medio da Lei percepito?*

	N	%	% cum
Meno di € 500	106	3,4	3,4
€ 500 - € 1000	275	8,7	12,1
€ 1001 - € 1500	1538	48,9	61,0
€ 1501 - € 2000	649	20,6	81,6
€ 2001 - € 2500	270	8,6	90,2
€ 2501 - € 3000	157	5,0	95,2
€ 3001 - € 3500	65	2,1	97,2
€ 3501 - € 4000	31	1,0	98,2
€ 4001 - € 6000	31	1,0	99,2
Oltre € 6000	25	0,8	100,0
<b>Totale validi</b>	<b>3147</b>	<b>100,0</b>	
Mancanti	341		
Totale assoluto	3488		

**Tabella C.A33**

*Negli ultimi tre mesi Lei ha cercato attivamente un nuovo lavoro?*

	N	%	% cum
Sì	954	29,1	29,1
No	2328	70,9	100,0
<b>Totale validi</b>	<b>3282</b>	<b>100,0</b>	
Mancanti	206		
Totale assoluto	3488		

**Tabella C.A34**

*Qual è la principale motivazione che l'ha indotta ad avviare la ricerca di un nuovo lavoro?*

	N	%	% cum
Stabilità del posto di lavoro	418	44,7	44,7
Insoddisfazione per il livello di retribuzione	124	13,2	57,9
Più elevata possibilità di carriera	113	12,1	70,0
Maggiore coerenza con gli studi fatti	56	6,0	76,0
Maggiore rispondenza al titolo posseduto	48	5,1	81,1
Possibilità di acquisire maggiore professionalità	65	6,9	88,0
Maggiore rispondenza ai propri interessi culturali	38	4,1	92,1
Maggiore indipendenza e autonomia sul lavoro	18	1,9	94,0
Localizzazione della sede di lavoro	49	5,2	99,3
Per avere più tempo libero	7	0,7	100,0
<b>Totale validi</b>	<b>936</b>	<b>100,0</b>	
Mancanti	18		
Totale assoluto	954		

## Blocco B: chi attualmente non lavora ma ha lavorato dopo il conseguimento del titolo

Tabella C.B01

L'ultimo lavoro svolto era lo stesso che svolgeva al momento del conseguimento del titolo? (Z07=1)

	N	%	% cum
Si	126	54,5	54,5
No	105	45,5	100,0
<b>Totale validi</b>	<b>231</b>	<b>100,0</b>	
Mancanti			
Totale assoluto			

Tabella C.B04

Il conseguimento del titolo di Dottore di Ricerca ha comportato un miglioramento nel suo lavoro? (B01=1)

	N	%	% cum
Si	33	28,7	28,7
No	82	71,3	100,0
<b>Totale validi</b>	<b>115</b>	<b>100,0</b>	
Mancanti			
Totale assoluto			

Tabella C.B04b.1

Livello di retribuzione attuale

	N	%	% cum
Si	10	35,7	35,7
No	18	64,3	100,0
<b>Totale validi</b>	<b>28</b>	<b>100,0</b>	
Mancanti	5		
Totale assoluto	33		

Tabella C.B04b.2

Prospettive di guadagno

	N	%	% cum
Si	13	48,1	48,1
No	14	51,9	100,0
<b>Totale validi</b>	<b>27</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	33		

Tabella C.B04b.3

Possibilità di carriera

	N	%	% cum
Si	17	60,7	60,7
No	11	39,3	100,0
<b>Totale validi</b>	<b>28</b>	<b>100,0</b>	
Mancanti	5		
Totale assoluto	33		

Tabella C.B04b.4

**Stabilità/sicurezza sul lavoro**

	N	%	% cum
Si	3	11,1	11,1
No	24	88,9	100,0
<b>Totale validi</b>	<b>27</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	33		

Tabella C.B04b.5

**Coerenza con studi fatti**

	N	%	% cum
Si	21	77,8	77,8
No	6	22,2	100,0
<b>Totale validi</b>	<b>27</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	33		

Tabella C.B04b.6

**Utilizzo delle competenze acquisite**

	N	%	% cum
Si	26	92,9	92,9
No	2	7,1	100,0
<b>Totale validi</b>	<b>28</b>	<b>100,0</b>	
Mancanti	5		
Totale assoluto	33		

Tabella C.B04b.7

**Acquisizione di professionalità**

	N	%	% cum
Si	25	86,2	86,2
No	4	13,8	100,0
<b>Totale validi</b>	<b>29</b>	<b>100,0</b>	
Mancanti	4		
Totale assoluto	33		

Tabella C.B04b.8

**Rispondenza ai propri interessi culturali**

	N	%	% cum
Si	26	92,9	92,9
No	2	7,1	100,0
<b>Totale validi</b>	<b>28</b>	<b>100,0</b>	
Mancanti	5		
Totale assoluto	33		

Tabella C.B04b.9

**Indipendenza e autonomia sul lavoro**

	N	%	% cum
Si	14	51,9	51,9
No	13	48,1	100,0
<b>Totale validi</b>	<b>27</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	33		

Tabella C.B04b.10

**Coinvolgimento nei processi decisionali**

	N	%	% cum
Si	17	60,7	60,7
No	11	39,3	100,0
<b>Totale validi</b>	<b>28</b>	<b>100,0</b>	
Mancanti	5		
Totale assoluto	33		

Tabella C.B04b.11

**Flessibilità dell'orario di lavoro**

	N	%	% cum
Si	7	25,9	25,9
No	20	74,1	100,0
<b>Totale validi</b>	<b>27</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	33		

Tabella C.B04b.12

**Tempo libero**

	N	%	% cum
Si	1	3,7	3,7
No	26	96,3	100,0
<b>Totale validi</b>	<b>27</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	33		

Tabella C.B04b.13

**Localizzazione sede di lavoro**

	N	%	% cum
Si	6	22,2	22,2
No	21	77,8	100,0
<b>Totale validi</b>	<b>27</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	33		

Tabella C.B04b.14

**Rapporti coi colleghi**

	N	%	% cum
Si	14	51,9	51,9
No	13	48,1	100,0
<b>Totale validi</b>	<b>27</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	33		

Tabella C.B04b.15

**Miglioramento generale rispetto alle aspettative iniziali**

	N	%	% cum
Si	16	59,3	59,3
No	11	40,7	100,0
<b>Totale validi</b>	<b>27</b>	<b>100,0</b>	
Mancanti	6		
Totale assoluto	33		

Tabella C.B04b.16

**Miglioramento rispetto alla attività lavorativa ideale**

	N	%	% cum
Sì	10	38,5	38,5
No	16	61,5	100,0
<b>Totale validi</b>	<b>26</b>	<b>100,0</b>	
Mancanti	7		
Totale assoluto	33		

Tabella C.B06

*Per quale motivo è stata interrotta la sua ultima attività lavorativa?*

	N	%	% cum
Per iniziare un'attività di formazione post-dottorato	41	9,2	9,2
Dimesso perché non soddisfatto del trattamento economico	14	3,1	12,4
Dimesso perché non soddisfatto di quel tipo di lavoro/contratto	26	5,8	18,2
Dimesso perché non soddisfatto della sede di lavoro	3	0,7	18,9
Dimesso per motivi personali	11	2,5	21,3
Licenziamento da parte del datore di lavoro o altra interruzione da lei <u>non voluta</u>	9	2,0	23,4
Scadenza regolare di un contratto a tempo determinato o termine di un lavoro occasionale o stagionale	222	49,9	73,3
Maternità	25	5,6	78,9
Altro	94	21,1	100,0
<b>Totale validi</b>	<b>445</b>	<b>100,0</b>	
Mancanti	97		
Totale assoluto	542		

Tabella C.B07

*Negli ultimi tre mesi Lei ha cercato attivamente un lavoro?*

	N	%	% cum
Sì	316	58,4	58,4
No	225	41,6	100,0
<b>Totale validi</b>	<b>541</b>	<b>100,0</b>	
Mancanti	1		
Totale assoluto	542		

## Blocco C: chi non ha mai lavorato dopo il conseguimento del titolo

*Tabella C.C01*

*Lei attualmente cerca lavoro?*

	N	%	% cum
Sì	99	81,1	81,1
No	23	18,9	100,0
<b>Totale validi</b>	<b>122</b>	<b>100,0</b>	
Mancanti			
Totale assoluto			

*Tabella C.C02*

*Quale tipo di lavoro cercava? (C01 = 1)*

	N	%	% cum
Dirigente	2	2,1	2,1
Direttivo/Quadro	3	3,2	5,3
Impiegato o intermedio	1	1,1	6,3
Ricercatore o Docente universitario	48	50,5	56,8
Ricercatore nel settore pubblico non universitario	6	6,3	63,2
Docente alle Scuole Medie (Inferiori o Superiori)	5	5,3	68,4
Lavoro autonomo	2	2,1	70,5
Libero professionista (iscritto albo)	0	0,0	70,5
Non ha preferenze	28	29,5	100,0
<b>Totale validi</b>	<b>95</b>	<b>100,0</b>	
Mancanti	4		
Totale assoluto	99		

*Tabella C.C04*

*Per quale motivo non cerca lavoro? (C01 = 2)*

	N	%	% cum
Motivi di studio o di formazione post-dottorato	13	56,5	56,5
Preparazione ad un concorso pubblico	2	8,7	65,2
Servizio civile nazionale volontario	0	0,0	65,2
In attesa di chiamata dopo aver superato prova / concorso / selezione	1	4,3	69,6
In attesa di avviare un'attività in proprio	2	8,7	78,3
Motivi personali (casalinga/o, maternità, assistenza figli o parent, problemi di salute)	5	21,7	100,0
Mancanza di opportunità di lavoro	0	0,0	100,0
Altri motivi	0	0,0	100,0
<b>Totale validi</b>	<b>23</b>	<b>100,0</b>	
Mancanti	0		
Totale assoluto	23		

## Tabelle per area disciplinare (aree Z, A, B e C)

**Tabella C.Z01a**

Qual è il motivo principale per cui ha deciso di iscriversi al Corso di Dottorato in cui ha conseguito il titolo di Dottore di Ricerca?

Valori assoluti

Area disciplinare	Per completare/ arricchire la mia formazione	Per conseguire un titolo dal più alto potere professionalizzante	Interesse per lo studio	Per fare attività di ricerca	Per l'interesse nei confronti di una futura carriera accademica	Altro	Totale
Ingegneria civile e Architettura	56	23	17	86	46	3	231
Ingegneria dell'informazione	38	18	3	61	20	6	146
Ingegneria industriale	53	21	10	81	35	10	210
Scienze agrarie	32	21	6	82	22	8	171
Scienze biologiche	99	61	27	248	58	7	500
Scienze chimiche	43	28	5	110	38	7	231
Scienze della terra	34	21	8	89	22	4	178
Scienze dell'antichità e filologico-letterarie	43	21	20	117	96	1	298
Scienze delle attività motorie e delle discipline sportive	3	5	0	8	4	0	20
Scienze economiche (aziendali)	8	13	6	20	28	1	76
Scienze economiche (socio-politiche)	23	17	10	41	37	3	131
Scienze fisiche	33	10	15	143	41	1	243
Scienze giuridiche	49	28	32	47	85	1	242
Scienze informatiche	2	2	1	13	3	0	21
Scienze matematiche	8	3	13	31	13	0	68
Scienze mediche	156	64	12	179	67	10	488
Scienze pedagogiche	9	6	2	18	5	0	40
Scienze politiche e sociali	27	12	13	41	30	3	126
Scienze psicologiche, geografiche e demoeoantropologiche	10	10	10	54	20	2	106
Scienze statistiche	5	4	4	4	5	2	24
Scienze storiche e filosofiche	19	9	23	79	60	2	192
Scienze storico-artistiche	10	3	7	42	17	1	80
Scienze veterinarie	12	7	2	23	9	3	56
<b>Totale</b>	<b>772</b>	<b>407</b>	<b>246</b>	<b>1617</b>	<b>761</b>	<b>75</b>	<b>3878</b>

**Tabella C.Z01b**

Qual è il motivo principale per cui ha deciso di iscriversi al Corso di Dottorato in cui ha conseguito il titolo di Dottore di Ricerca?

Valori percentuali di riga

Area disciplinare	Per completare/ arricchire la mia formazione	Per conseguire un titolo dal più alto potere professionalizzante	Interesse per lo studio	Per fare attività di ricerca	Per l'interesse nei confronti di una futura carriera accademica	Altro	Totale
Ingegneria civile e Architettura	24,2	10,0	7,4	37,2	19,9	1,3	100,0
Ingegneria dell'informazione	26,0	12,3	2,1	41,8	13,7	4,1	100,0
Ingegneria industriale	25,2	10,0	4,8	38,6	16,7	4,8	100,0
Scienze agrarie	18,7	12,3	3,5	48,0	12,9	4,7	100,0
Scienze biologiche	19,8	12,2	5,4	49,6	11,6	1,4	100,0
Scienze chimiche	18,6	12,1	2,2	47,6	16,5	3,0	100,0
Scienze della terra	19,1	11,8	4,5	50,0	12,4	2,2	100,0
Scienze dell'antichità e filologico-letterarie	14,4	7,0	6,7	39,3	32,2	0,3	100,0
Scienze delle attività motorie e delle discipline sportive	15,0	25,0	0,0	40,0	20,0	0,0	100,0
Scienze economiche (aziendali)	10,5	17,1	7,9	26,3	36,8	1,3	100,0
Scienze economiche (socio-politiche)	17,6	13,0	7,6	31,3	28,2	2,3	100,0
Scienze fisiche	13,6	4,1	6,2	58,8	16,9	0,4	100,0
Scienze giuridiche	20,2	11,6	13,2	19,4	35,1	0,4	100,0
Scienze informatiche	9,5	9,5	4,8	61,9	14,3	0,0	100,0
Scienze matematiche	11,8	4,4	19,1	45,6	19,1	0,0	100,0
Scienze mediche	32,0	13,1	2,5	36,7	13,7	2,0	100,0
Scienze pedagogiche	22,5	15,0	5,0	45,0	12,5	0,0	100,0
Scienze politiche e sociali	21,4	9,5	10,3	32,5	23,8	2,4	100,0
Scienze psicologiche, geografiche e demoeoantropologiche	9,4	9,4	9,4	50,9	18,9	1,9	100,0
Scienze statistiche	20,8	16,7	16,7	16,7	20,8	8,3	100,0
Scienze storiche e filosofiche	9,9	4,7	12,0	41,1	31,3	1,0	100,0
Scienze storico-artistiche	12,5	3,8	8,8	52,5	21,3	1,3	100,0
Scienze veterinarie	21,4	12,5	3,6	41,1	16,1	5,4	100,0
<b>Totale</b>	<b>19,9</b>	<b>10,5</b>	<b>6,3</b>	<b>41,7</b>	<b>19,6</b>	<b>1,9</b>	<b>100,0</b>



Tabella C.Z04a

Se potesse tornare indietro, si iscriverebbe di nuovo ad un Corso di Dottorato?

Valori assoluti

Area disciplinare	Sì, allo stesso Corso nella stessa sede universitaria	Sì, allo stesso Corso ma in altra sede universitaria	Sì ma ad altro Corso	No	Totale
Ingegneria civile e Architettura	117	55	28	27	227
Ingegneria dell'informazione	86	22	5	27	140
Ingegneria industriale	129	24	17	34	204
Scienze agrarie	80	25	27	38	170
Scienze biologiche	223	82	85	98	488
Scienze chimiche	127	34	20	48	229
Scienze della terra	85	33	25	33	176
Scienze dell'antichità e filologico-letterarie	160	63	25	44	292
Scienze delle attività motorie e delle discipline sportive	5	2	5	8	20
Scienze economiche (aziendali)	39	20	10	5	74
Scienze economiche (socio-politiche)	60	38	17	15	130
Scienze fisiche	140	52	13	32	237
Scienze giuridiche	152	43	24	22	241
Scienze informatiche	11	8	0	2	21
Scienze matematiche	39	13	3	11	66
Scienze mediche	253	53	70	99	475
Scienze pedagogiche	25	7	4	4	40
Scienze politiche e sociali	55	32	18	17	122
Scienze psicologiche, geografiche e demoeoantropologiche	51	23	10	20	104
Scienze statistiche	12	7	1	3	23
Scienze storiche e filosofiche	99	48	14	27	188
Scienze storico-artistiche	42	20	9	9	80
Scienze veterinarie	30	6	11	10	57
<b>Totale</b>	<b>2020</b>	<b>710</b>	<b>441</b>	<b>633</b>	<b>3804</b>

Tabella C.Z04b

Se potesse tornare indietro, si iscriverebbe di nuovo ad un Corso di Dottorato?

Valori percentuali di riga

Area disciplinare	Sì, allo stesso Corso nella stessa sede universitaria	Sì, allo stesso Corso ma in altra sede universitaria	Sì ma ad altro Corso	No	Totale
Ingegneria civile e Architettura	51,5	24,2	12,3	11,9	100,0
Ingegneria dell'informazione	61,4	15,7	3,6	19,3	100,0
Ingegneria industriale	63,2	11,8	8,3	16,7	100,0
Scienze agrarie	47,1	14,7	15,9	22,4	100,0
Scienze biologiche	45,7	16,8	17,4	20,1	100,0
Scienze chimiche	55,5	14,8	8,7	21,0	100,0
Scienze della terra	48,3	18,8	14,2	18,8	100,0
Scienze dell'antichità e filologico-letterarie	54,8	21,6	8,6	15,1	100,0
Scienze delle attività motorie e delle discipline sportive	25,0	10,0	25,0	40,0	100,0
Scienze economiche (aziendali)	52,7	27,0	13,5	6,8	100,0
Scienze economiche (socio-politiche)	46,2	29,2	13,1	11,5	100,0
Scienze fisiche	59,1	21,9	5,5	13,5	100,0
Scienze giuridiche	63,1	17,8	10,0	9,1	100,0
Scienze informatiche	52,4	38,1		9,5	100,0
Scienze matematiche	59,1	19,7	4,5	16,7	100,0
Scienze mediche	53,3	11,2	14,7	20,8	100,0
Scienze pedagogiche	62,5	17,5	10,0	10,0	100,0
Scienze politiche e sociali	45,1	26,2	14,8	13,9	100,0
Scienze psicologiche, geografiche e demoeoantropologiche	49,0	22,1	9,6	19,2	100,0
Scienze statistiche	52,2	30,4	4,3	13,0	100,0
Scienze storiche e filosofiche	52,7	25,5	7,4	14,4	100,0
Scienze storico-artistiche	52,5	25,0	11,3	11,3	100,0
Scienze veterinarie	52,6	10,5	19,3	17,5	100,0
<b>Totale</b>	<b>53,1</b>	<b>18,7</b>	<b>11,6</b>	<b>16,6</b>	<b>100,0</b>

Tabella C.Z06a

Se potesse tornare indietro, si iscriverebbe di nuovo ad un Corso di Dottorato?

Valori assoluti

Area disciplinare	Sì, attività attinente all'attività di ricerca del dottorato	Sì, attività non attinente all'attività di ricerca del dotto	No	Totale
Ingegneria civile e Architettura	91	79	62	232
Ingegneria dell'informazione	71	20	54	145
Ingegneria industriale	93	47	68	208
Scienze agrarie	47	40	84	171
Scienze biologiche	152	72	276	500
Scienze chimiche	67	41	123	231
Scienze della terra	72	28	78	178
Scienze dell'antichità e filologico-letterarie	76	103	120	299
Scienze delle attività motorie e delle discipline sportive	4	6	10	20
Scienze economiche (aziendali)	44	20	12	76
Scienze economiche (socio-politiche)	67	31	31	129
Scienze fisiche	75	39	129	243
Scienze giuridiche	90	69	85	244
Scienze informatiche	9	6	6	21
Scienze matematiche	22	15	31	68
Scienze mediche	177	102	205	484
Scienze pedagogiche	20	8	12	40
Scienze politiche e sociali	50	49	28	127
Scienze psicologiche, geografiche e demografiche e antropologiche	31	32	43	106
Scienze statistiche	12	2	9	23
Scienze storiche e filosofiche	44	78	69	191
Scienze storico-artistiche	21	30	29	80
Scienze veterinarie	16	13	27	56
<b>Totale</b>	<b>1351</b>	<b>930</b>	<b>1591</b>	<b>3872</b>

Tabella C.Z06b

Se potesse tornare indietro, si iscriverebbe di nuovo ad un Corso di Dottorato?

Valori percentuali di riga

Area disciplinare	Sì, attività attinente all'attività di ricerca del dottorato	Sì, attività non attinente all'attività di ricerca del dotto	No	Totale
Ingegneria civile e Architettura	39,2	34,1	26,7	100,0
Ingegneria dell'informazione	49,0	13,8	37,2	100,0
Ingegneria industriale	44,7	22,6	32,7	100,0
Scienze agrarie	27,5	23,4	49,1	100,0
Scienze biologiche	30,4	14,4	55,2	100,0
Scienze chimiche	29,0	17,7	53,2	100,0
Scienze della terra	40,4	15,7	43,8	100,0
Scienze dell'antichità e filologico-letterarie	25,4	34,4	40,1	100,0
Scienze delle attività motorie e delle discipline sportive	20,0	30,0	50,0	100,0
Scienze economiche (aziendali)	57,9	26,3	15,8	100,0
Scienze economiche (socio-politiche)	51,9	24,0	24,0	100,0
Scienze fisiche	30,9	16,0	53,1	100,0
Scienze giuridiche	36,9	28,3	34,8	100,0
Scienze informatiche	42,9	28,6	28,6	100,0
Scienze matematiche	32,4	22,1	45,6	100,0
Scienze mediche	36,6	21,1	42,4	100,0
Scienze pedagogiche	50,0	20,0	30,0	100,0
Scienze politiche e sociali	39,4	38,6	22,0	100,0
Scienze psicologiche, geografiche e demografiche e antropologiche	29,2	30,2	40,6	100,0
Scienze statistiche	52,2	8,7	39,1	100,0
Scienze storiche e filosofiche	23,0	40,8	36,1	100,0
Scienze storico-artistiche	26,3	37,5	36,3	100,0
Scienze veterinarie	28,6	23,2	48,2	100,0
<b>Totale</b>	<b>34,9</b>	<b>24,0</b>	<b>41,1</b>	<b>100,0</b>

Tabella C.Z05a

Durante il Corso di Dottorato ha compiuto periodi di studio all'estero?

Valori assoluti

Area disciplinare	Sì	No	Totale
Ingegneria civile e Architettura	77	154	231
Ingegneria dell'informazione	64	80	144
Ingegneria industriale	84	123	207
Scienze agrarie	84	87	171
Scienze biologiche	160	335	495
Scienze chimiche	94	135	229
Scienze della terra	66	111	177
Scienze dell'antichità e filologico-letterarie	148	147	295
Scienze delle attività motorie e delle discipline sportive	4	16	20
Scienze economiche (aziendali)	37	40	77
Scienze economiche (socio-politiche)	72	57	129
Scienze fisiche	106	133	239
Scienze giuridiche	94	147	241
Scienze informatiche	7	13	20
Scienze matematiche	33	35	68
Scienze mediche	112	372	484
Scienze pedagogiche	13	27	40
Scienze politiche e sociali	67	59	126
Scienze psicologiche, geografiche e demografiche e demografiche	55	50	105
Scienze statistiche	16	7	23
Scienze storiche e filosofiche	94	95	189
Scienze storico-artistiche	18	60	78
Scienze veterinarie	28	29	57
<b>Totale</b>	<b>1533</b>	<b>2312</b>	<b>3845</b>

Tabella C.Z05b

Durante il Corso di Dottorato ha compiuto periodi di studio all'estero?

Valori percentuali di riga

Area disciplinare	Sì	No	Totale
Ingegneria civile e Architettura	33,3	66,7	100,0
Ingegneria dell'informazione	44,4	55,6	100,0
Ingegneria industriale	40,6	59,4	100,0
Scienze agrarie	49,1	50,9	100,0
Scienze biologiche	32,3	67,7	100,0
Scienze chimiche	41,0	59,0	100,0
Scienze della terra	37,3	62,7	100,0
Scienze dell'antichità e filologico-letterarie	50,2	49,8	100,0
Scienze delle attività motorie e delle discipline sportive	20,0	80,0	100,0
Scienze economiche (aziendali)	48,1	51,9	100,0
Scienze economiche (socio-politiche)	55,8	44,2	100,0
Scienze fisiche	44,4	55,6	100,0
Scienze giuridiche	39,0	61,0	100,0
Scienze informatiche	35,0	65,0	100,0
Scienze matematiche	48,5	51,5	100,0
Scienze mediche	23,1	76,9	100,0
Scienze pedagogiche	32,5	67,5	100,0
Scienze politiche e sociali	53,2	46,8	100,0
Scienze psicologiche, geografiche e demografiche e demografiche	52,4	47,6	100,0
Scienze statistiche	69,6	30,4	100,0
Scienze storiche e filosofiche	49,7	50,3	100,0
Scienze storico-artistiche	23,1	76,9	100,0
Scienze veterinarie	49,1	50,9	100,0
<b>Totale</b>	<b>39,9</b>	<b>60,1</b>	<b>100,0</b>

Tabella C.207a

Lei lavorava al momento del conseguimento del titolo di Dottore di Ricerca?

Valori assoluti

Area disciplinare	Sì	No	Totale
Ingegneria civile e Architettura	124	110	234
Ingegneria dell'informazione	63	83	146
Ingegneria industriale	107	103	210
Scienze agrarie	73	101	174
Scienze biologiche	221	282	503
Scienze chimiche	88	143	231
Scienze della terra	84	95	179
Scienze dell'antichità e filologico-letterarie	135	167	302
Scienze delle attività motorie e delle discipline sportive	8	12	20
Scienze economiche (aziendali)	45	32	77
Scienze economiche (socio-politiche)	66	66	132
Scienze fisiche	87	160	247
Scienze giuridiche	130	118	248
Scienze informatiche	12	9	21
Scienze matematiche	22	46	68
Scienze mediche	289	200	489
Scienze pedagogiche	28	12	40
Scienze politiche e sociali	69	58	127
Scienze psicologiche, geografiche e demograficoantropologiche	51	55	106
Scienze statistiche	6	18	24
Scienze storiche e filosofiche	83	110	193
Scienze storico-artistiche	32	48	80
Scienze veterinarie	26	31	57
<b>Totale</b>	<b>1849</b>	<b>2059</b>	<b>3908</b>

Tabella C.207b

Lei lavorava al momento del conseguimento del titolo di Dottore di Ricerca?

Valori percentuali di riga

Area disciplinare	Sì	No	Totale
Ingegneria civile e Architettura	53,0	47,0	100,0
Ingegneria dell'informazione	43,2	56,8	100,0
Ingegneria industriale	51,0	49,0	100,0
Scienze agrarie	42,0	58,0	100,0
Scienze biologiche	43,9	56,1	100,0
Scienze chimiche	38,1	61,9	100,0
Scienze della terra	46,9	53,1	100,0
Scienze dell'antichità e filologico-letterarie	44,7	55,3	100,0
Scienze delle attività motorie e delle discipline sportive	40,0	60,0	100,0
Scienze economiche (aziendali)	58,4	41,6	100,0
Scienze economiche (socio-politiche)	50,0	50,0	100,0
Scienze fisiche	35,2	64,8	100,0
Scienze giuridiche	52,4	47,6	100,0
Scienze informatiche	57,1	42,9	100,0
Scienze matematiche	32,4	67,6	100,0
Scienze mediche	59,1	40,9	100,0
Scienze pedagogiche	70,0	30,0	100,0
Scienze politiche e sociali	54,3	45,7	100,0
Scienze psicologiche, geografiche e demograficoantropologiche	48,1	51,9	100,0
Scienze statistiche	25,0	75,0	100,0
Scienze storiche e filosofiche	43,0	57,0	100,0
Scienze storico-artistiche	40,0	60,0	100,0
Scienze veterinarie	45,6	54,4	100,0
<b>Totale</b>	<b>47,3</b>	<b>52,7</b>	<b>100,0</b>

Tabella C.Z08a

Indichi la sua condizione occupazionale attuale  
Valori assoluti

Area disciplinare	Lavoro con assegno di ricerca, borsa post-doc o altra borsa	Lavoro con altra forma contrattuale a tempo determinato	Lavoro a tempo indeterminato/ autonomo effettivo	Attività di formazione post-dottorato retribuita	Attività di formazione post-dottorato non retribuita	Periodo di studio in preparazione a concorsi pubblici/esame	Ricerca di un lavoro	Altra condizione di inattività (non lavora e non cerca lavoro)	Altro	Totale
Ingegneria civile e Architettura	70	42	81	2	1	2	12	2	19	231
Ingegneria dell'informazione	56	24	53	6	1	0	2	0	4	146
Ingegneria industriale	74	25	88	2	3	0	3	0	12	207
Scienze agrarie	86	28	31	3	2	0	11	0	10	171
Scienze biologiche	221	100	99	11	7	4	26	2	31	501
Scienze chimiche	90	49	64	10	3	1	5	1	6	229
Scienze della terra	66	40	44	2	1	0	12	1	9	175
Scienze dell'antichità e filologico-letterarie	64	102	56	5	9	7	36	1	18	298
Scienze delle attività motorie e delle discipline sportive	11	5	2	0	0	0	2	0	0	20
Scienze economiche (aziendali)	17	14	36	0	0	2	2	0	4	75
Scienze economiche (socio-politiche)	46	34	37	2	1	0	7	0	3	130
Scienze fisiche	142	41	46	4	2	1	7	1	3	247
Scienze giuridiche	53	35	103	5	8	8	11	2	19	244
Scienze informatiche	5	8	5	0	0	0	0	0	2	20
Scienze matematiche	41	8	10	0	0	0	4	0	4	67
Scienze mediche	160	114	138	12	8	3	22	4	23	484
Scienze pedagogiche	10	6	17	1	1	1	1	0	2	39
Scienze politiche e sociali	55	34	26	0	2	4	4	0	2	127
Scienze psicologiche, geografiche e demotnoantropologiche	39	25	28	1	0	2	3	1	5	104
Scienze statistiche	10	5	7	0	0	0	2	0	0	24
Scienze storiche e filosofiche	49	54	38	5	5	5	26	1	8	191
Scienze storico-artistiche	17	19	24	0	4	5	7	0	4	80
Scienze veterinarie	20	11	13	0	1	0	4	2	4	55
<b>Totale</b>	<b>1402</b>	<b>823</b>	<b>1046</b>	<b>71</b>	<b>59</b>	<b>45</b>	<b>209</b>	<b>18</b>	<b>192</b>	<b>3865</b>

Tabella C.Z08b

Indichi la sua condizione occupazionale attuale  
Valori percentuali di riga

Area disciplinare	Lavoro con assegno di ricerca, borsa post-doc o altra borsa	Lavoro con altra forma contrattuale a tempo determinato	Lavoro a tempo indeterminato/ autonomo effettivo	Attività di formazione post-dottorato retribuita	Attività di formazione post-dottorato non retribuita	Periodo di studio in preparazione a concorsi pubblici/esame	Ricerca di un lavoro	Altra condizione di inattività (non lavora e non cerca lavoro)	Altro	Totale
Ingegneria civile e Architettura	30,3	18,2	35,1	0,9	0,4	0,9	5,2	0,9	8,2	100,0
Ingegneria dell'informazione	38,4	16,4	36,3	4,1	0,7	0,0	1,4	0,0	2,7	100,0
Ingegneria industriale	35,7	12,1	42,5	1,0	1,4	0,0	1,4	0,0	5,8	100,0
Scienze agrarie	50,3	16,4	18,1	1,8	1,2	0,0	6,4	0,0	5,8	100,0
Scienze biologiche	44,1	20,0	19,8	2,2	1,4	0,8	5,2	0,4	6,2	100,0
Scienze chimiche	39,3	21,4	27,9	4,4	1,3	0,4	2,2	0,4	2,6	100,0
Scienze della terra	37,7	22,9	25,1	1,1	0,6	0,0	6,9	0,6	5,1	100,0
Scienze dell'antichità e filologico-letterarie	21,5	34,2	18,8	1,7	3,0	2,3	12,1	0,3	6,0	100,0
Scienze delle attività motorie e delle discipline sportive	55,0	25,0	10,0	0,0	0,0	0,0	10,0	0,0	0,0	100,0
Scienze economiche (aziendali)	22,7	18,7	48,0	0,0	0,0	2,7	2,7	0,0	5,3	100,0
Scienze economiche (socio-politiche)	35,4	26,2	28,5	1,5	0,8	0,0	5,4	0,0	2,3	100,0
Scienze fisiche	57,5	16,6	18,6	1,6	0,8	0,4	2,8	0,4	1,2	100,0
Scienze giuridiche	21,7	14,3	42,2	2,0	3,3	3,3	4,5	0,8	7,8	100,0
Scienze informatiche	25,0	40,0	25,0	0,0	0,0	0,0	0,0	0,0	10,0	100,0
Scienze matematiche	61,2	11,9	14,9	0,0	0,0	0,0	6,0	0,0	6,0	100,0
Scienze mediche	33,1	23,6	28,5	2,5	1,7	0,6	4,5	0,8	4,8	100,0
Scienze pedagogiche	25,6	15,4	43,6	2,6	2,6	2,6	2,6	0,0	5,1	100,0
Scienze politiche e sociali	43,3	26,8	20,5	0,0	1,6	3,1	3,1	0,0	1,6	100,0
Scienze psicologiche, geografiche e demotnoantropologiche	37,5	24,0	26,9	1,0	0,0	1,9	2,9	1,0	4,8	100,0
Scienze statistiche	41,7	20,8	29,2	0,0	0,0	0,0	8,3	0,0	0,0	100,0
Scienze storiche e filosofiche	25,7	28,3	19,9	2,6	2,6	2,6	13,6	0,5	4,2	100,0
Scienze storico-artistiche	21,3	23,8	30,0	0,0	5,0	6,3	8,8	0,0	5,0	100,0
Scienze veterinarie	36,4	20,0	23,6	0,0	1,8	0,0	7,3	3,6	7,3	100,0
<b>Totale</b>	<b>36,3</b>	<b>21,3</b>	<b>27,1</b>	<b>1,8</b>	<b>1,5</b>	<b>1,2</b>	<b>5,4</b>	<b>0,5</b>	<b>5,0</b>	<b>100,0</b>

Tabella C.A011a

Il suo attuale lavoro è di tipo autonomo o dipendente?

Valori assoluti

Area disciplinare	Autonomo	Dipendente inclusi contratti di collaborazione	Totale
Ingegneria civile e Architettura	61	127	188
Ingegneria dell'informazione	13	118	131
Ingegneria industriale	19	165	184
Scienze agrarie	10	135	145
Scienze biologiche	24	389	413
Scienze chimiche	9	194	203
Scienze della terra	11	137	148
Scienze dell'antichità e filologico-letterarie	17	204	221
Scienze delle attività motorie e delle discipline sportive	1	17	18
Scienze economiche (aziendali)	18	47	65
Scienze economiche (socio-politiche)	15	101	116
Scienze fisiche	11	216	227
Scienze giuridiche	75	115	190
Scienze informatiche	0	18	18
Scienze matematiche	2	57	59
Scienze mediche	61	345	406
Scienze pedagogiche	4	29	33
Scienze politiche e sociali	15	100	115
Scienze psicologiche, geografiche e demotnoantropologiche	18	75	93
Scienze statistiche	3	19	22
Scienze storiche e filosofiche	10	130	140
Scienze storico-artistiche	8	51	59
Scienze veterinarie	5	39	44
<b>Totale</b>	<b>410</b>	<b>2828</b>	<b>3238</b>

Tabella C.A011a

Il suo attuale lavoro è di tipo autonomo o dipendente?

Valori percentuali di riga

Area disciplinare	Autonomo	Dipendente inclusi contratti di collaborazione	Totale
Ingegneria civile e Architettura	32,4	67,6	100,0
Ingegneria dell'informazione	9,9	90,1	100,0
Ingegneria industriale	10,3	89,7	100,0
Scienze agrarie	6,9	93,1	100,0
Scienze biologiche	5,8	94,2	100,0
Scienze chimiche	4,4	95,6	100,0
Scienze della terra	7,4	92,6	100,0
Scienze dell'antichità e filologico-letterarie	7,7	92,3	100,0
Scienze delle attività motorie e delle discipline sportive	5,6	94,4	100,0
Scienze economiche (aziendali)	27,7	72,3	100,0
Scienze economiche (socio-politiche)	12,9	87,1	100,0
Scienze fisiche	4,8	95,2	100,0
Scienze giuridiche	39,5	60,5	100,0
Scienze informatiche	0,0	100,0	100,0
Scienze matematiche	3,4	96,6	100,0
Scienze mediche	15,0	85,0	100,0
Scienze pedagogiche	12,1	87,9	100,0
Scienze politiche e sociali	13,0	87,0	100,0
Scienze psicologiche, geografiche e demotnoantropologiche	19,4	80,6	100,0
Scienze statistiche	13,6	86,4	100,0
Scienze storiche e filosofiche	7,1	92,9	100,0
Scienze storico-artistiche	13,6	86,4	100,0
Scienze veterinarie	11,4	88,6	100,0
<b>Totale</b>	<b>12,7</b>	<b>87,3</b>	<b>100,0</b>

Tabella C.A016a

Nel suo lavoro svolge mansioni di ricerca e sviluppo? (A14<=3 o A11=1)

Valori assoluti

Area disciplinare	Sì, attività attinente all'attività di ricerca del dottorato	Sì, attività non attinente all'attività di ricerca del dotto	No	Totale
Ingegneria civile e Architettura	3	12	25	40
Ingegneria dell'informazione	10	13	15	38
Ingegneria industriale	16	13	22	51
Scienze agrarie	1	6	12	19
Scienze biologiche	12	24	33	69
Scienze chimiche	12	16	27	55
Scienze della terra	7	14	12	33
Scienze dell'antichità e filologico-letterarie	2	16	16	34
Scienze delle attività motorie e delle discipline sportive	1	1	0	2
Scienze economiche (aziendali)	1	10	7	18
Scienze economiche (socio-politiche)	2	14	11	27
Scienze fisiche	4	10	18	32
Scienze giuridiche	3	19	20	42
Scienze informatiche	2	1	3	6
Scienze matematiche	1	1	3	5
Scienze mediche	10	80	28	118
Scienze pedagogiche	0	1	0	1
Scienze politiche e sociali	1	14	11	26
Scienze psicologiche, geografiche e demografiche e demografiche e antropologiche	1	3	5	9
Scienze statistiche	2	3	1	6
Scienze storiche e filosofiche	2	13	21	36
Scienze storico-artistiche	1	7	5	13
Scienze veterinarie	2	3	1	6
<b>Totale</b>	<b>96</b>	<b>294</b>	<b>296</b>	<b>686</b>

Tabella C.A016b

Nel suo lavoro svolge mansioni di ricerca e sviluppo? (A14<=3 o A11=1)

Valori percentuali di riga

Area disciplinare	Sì, attività attinente all'attività di ricerca del dottorato	Sì, attività non attinente all'attività di ricerca del dotto	No	Totale
Ingegneria civile e Architettura	7,5	30,0	62,5	100,0
Ingegneria dell'informazione	26,3	34,2	39,5	100,0
Ingegneria industriale	31,4	25,5	43,1	100,0
Scienze agrarie	5,3	31,6	63,2	100,0
Scienze biologiche	17,4	34,8	47,8	100,0
Scienze chimiche	21,8	29,1	49,1	100,0
Scienze della terra	21,2	42,4	36,4	100,0
Scienze dell'antichità e filologico-letterarie	5,9	47,1	47,1	100,0
Scienze delle attività motorie e delle discipline sportive	50,0	50,0	0,0	100,0
Scienze economiche (aziendali)	5,6	55,6	38,9	100,0
Scienze economiche (socio-politiche)	7,4	51,9	40,7	100,0
Scienze fisiche	12,5	31,3	56,3	100,0
Scienze giuridiche	7,1	45,2	47,6	100,0
Scienze informatiche	33,3	16,7	50,0	100,0
Scienze matematiche	20,0	20,0	60,0	100,0
Scienze mediche	8,5	67,8	23,7	100,0
Scienze pedagogiche	0,0	100,0	0,0	100,0
Scienze politiche e sociali	3,8	53,8	42,3	100,0
Scienze psicologiche, geografiche e demografiche e demografiche e antropologiche	11,1	33,3	55,6	100,0
Scienze statistiche	33,3	50,0	16,7	100,0
Scienze storiche e filosofiche	5,6	36,1	58,3	100,0
Scienze storico-artistiche	7,7	53,8	38,5	100,0
Scienze veterinarie	33,3	50,0	16,7	100,0
<b>Totale</b>	<b>14,0</b>	<b>42,9</b>	<b>43,1</b>	<b>100,0</b>

Tabella C.A017a

Nel suo lavoro svolge attività di innovazione? (A14<=3 o A11=1)

Valori assoluti

Area disciplinare	Si, sia innovazione di processo che innovazione di prodotto	Si, innovazione di processo	Si, innovazione di prodotto	No	Totale
Ingegneria civile e Architettura	8	3	4	25	40
Ingegneria dell'informazione	6	2	16	14	38
Ingegneria industriale	18	8	9	16	51
Scienze agrarie	2	3	0	14	19
Scienze biologiche	13	17	5	34	69
Scienze chimiche	16	7	4	28	55
Scienze della terra	6	5	2	20	33
Scienze dell'antichità e filologico-letterarie	5	3	2	23	33
Scienze delle attività motorie e delle discipline sportive	1	0	0	1	2
Scienze economiche (aziendali)	2	7	1	8	18
Scienze economiche (socio-politiche)	3	5	2	16	26
Scienze fisiche	11	4	7	10	32
Scienze giuridiche	2	8	0	30	40
Scienze informatiche	2	1	0	3	6
Scienze matematiche	1	2	0	2	5
Scienze mediche	20	21	6	66	113
Scienze pedagogiche	0	0	1	0	1
Scienze politiche e sociali	4	4	1	16	25
Scienze psicologiche, geografiche e demoeoantropologiche	2	3	0	3	8
Scienze statistiche	2	3	0	1	6
Scienze storiche e filosofiche	3	4	2	26	35
Scienze storico-artistiche	2	6	1	4	13
Scienze veterinarie	3	1	0	2	6
<b>Totale</b>	<b>132</b>	<b>117</b>	<b>63</b>	<b>362</b>	<b>674</b>

Tabella C.A017b

Nel suo lavoro svolge attività di innovazione? (A14<=3 o A11=1)

Valori percentuali di riga

Area disciplinare	Si, sia innovazione di processo che innovazione di prodotto	Si, innovazione di processo	Si, innovazione di prodotto	No	Totale
Ingegneria civile e Architettura	20,0	7,5	10,0	62,5	100,0
Ingegneria dell'informazione	15,8	5,3	42,1	36,8	100,0
Ingegneria industriale	35,3	15,7	17,6	31,4	100,0
Scienze agrarie	10,5	15,8	0,0	73,7	100,0
Scienze biologiche	18,8	24,6	7,2	49,3	100,0
Scienze chimiche	29,1	12,7	7,3	50,9	100,0
Scienze della terra	18,2	15,2	6,1	60,6	100,0
Scienze dell'antichità e filologico-letterarie	15,2	9,1	6,1	69,7	100,0
Scienze delle attività motorie e delle discipline sportive	50,0	0,0	0,0	50,0	100,0
Scienze economiche (aziendali)	11,1	38,9	5,6	44,4	100,0
Scienze economiche (socio-politiche)	11,5	19,2	7,7	61,5	100,0
Scienze fisiche	34,4	12,5	21,9	31,3	100,0
Scienze giuridiche	5,0	20,0	0,0	75,0	100,0
Scienze informatiche	33,3	16,7	0,0	50,0	100,0
Scienze matematiche	20,0	40,0	0,0	40,0	100,0
Scienze mediche	17,7	18,6	5,3	58,4	100,0
Scienze pedagogiche	0,0	0,0	100,0	0,0	100,0
Scienze politiche e sociali	16,0	16,0	4,0	64,0	100,0
Scienze psicologiche, geografiche e demoeoantropologiche	25,0	37,5	0,0	37,5	100,0
Scienze statistiche	33,3	50,0	0,0	16,7	100,0
Scienze storiche e filosofiche	8,6	11,4	5,7	74,3	100,0
Scienze storico-artistiche	15,4	46,2	7,7	30,8	100,0
Scienze veterinarie	50,0	16,7	0,0	33,3	100,0
<b>Totale</b>	<b>19,6</b>	<b>17,4</b>	<b>9,3</b>	<b>53,7</b>	<b>100,0</b>



Tabella C.A014a

Può qualificarmi sotto il profilo giuridico il suo lavoro attuale?

Valori assoluti

Area disciplinare	Dirigente	Direttivo/quadro	Impiegato o intermedio	Ricercatore nel settore privato	Assegnista/bo rsista	Ricercatore universitario a tempo determinato	Ricercatore universitario di ruolo	Professore universitario associato	Professore universitario ordinario	Ricercatore (settore pubblico non universitario)	Primo ricercatore (settore pubblico non universitario)	Dirigente di ricerca (settore pubblico non universitario)	Docente alle scuole medie inferiori	Docente alle scuole medie superiori	Altro	Totale
Ingegneria civile e Architettura	1	4	35	2	48	4	2	0	1	5	0	0	4	6	14	126
Ingegneria dell'informazione	0	6	32	8	44	6	4	0	0	7	0	0	1	1	6	115
Ingegneria industriale	3	5	43	11	58	9	13	0	1	10	1	0	0	4	6	164
Scienze agrarie	1	2	16	7	69	9	2	0	0	14	1	0	2	2	9	134
Scienze biologiche	15	6	48	23	186	20	5	0	1	31	0	2	8	5	37	387
Scienze chimiche	8	4	43	21	71	11	5	0	0	9	0	0	1	2	18	193
Scienze della terra	0	7	28	5	55	5	3	0	0	13	1	1	5	5	9	137
Scienze dell'antichità e filologico-letterarie	1	5	27	4	53	9	6	2	0	2	0	0	15	39	35	198
Scienze delle attività motorie e delle discipline sportive	2	0	0	0	10	0	0	0	0	0	0	0	1	0	3	16
Scienze economiche (aziendali)	1	5	12	1	12	3	8	2	0	0	0	0	0	1	2	47
Scienze economiche (socio-politiche)	2	6	19	5	39	7	4	2	0	4	0	0	1	0	10	99
Scienze fisiche	1	3	29	6	108	24	5	0	0	23	0	0	1	8	7	215
Scienze giuridiche	5	18	20	2	34	3	9	1	1	3	1	0	2	1	13	113
Scienze informatiche	0	2	4	2	3	2	1	0	0	4	0	0	0	0	0	18
Scienze matematiche	1	0	4	2	35	4	2	0	0	2	0	0	2	3	2	57
Scienze mediche	94	6	18	13	126	13	12	0	0	21	1	0	4	7	27	342
Scienze pedagogiche	1	0	0	0	9	1	1	0	0	0	0	0	0	6	11	29
Scienze politiche e sociali	2	5	18	6	40	6	4	0	0	2	0	1	0	1	14	99
Scienze psicologiche, geografiche e demotnoantropologiche	1	0	8	3	35	5	5	0	0	1	0	0	0	4	12	74
Scienze statistiche	0	2	4	0	9	1	1	0	0	0	0	0	0	0	2	19
Scienze storiche e filosofiche	1	6	26	2	39	4	0	1	0	2	0	0	8	19	19	127
Scienze storico-artistiche	0	4	9	1	15	2	1	0	0	0	0	0	3	6	10	51
Scienze veterinarie	3	1	2	2	19	3	3	0	0	1	0	0	0	0	4	38
<b>Totale</b>	<b>143</b>	<b>97</b>	<b>445</b>	<b>126</b>	<b>1117</b>	<b>151</b>	<b>96</b>	<b>8</b>	<b>4</b>	<b>154</b>	<b>5</b>	<b>4</b>	<b>58</b>	<b>120</b>	<b>270</b>	<b>2798</b>

Tabella C.A014b

Può qualificarmi sotto il profilo giuridico il suo lavoro attuale?

Valori percentuali di riga

Area disciplinare	Dirigente	Direttivo/quadro	Impiegato o intermedio	Ricercatore nel settore privato	Assegnista/bo-rsista	Ricercatore universitario a tempo determinato	Ricercatore universitario di ruolo	Professore universitario associato	Professore universitario ordinario	Ricercatore (settore pubblico non universitario)	Primo ricercatore (settore pubblico non universitario)	Dirigente di ricerca (settore pubblico non universitario)	Docente alle scuole medie inferiori	Docente alle scuole medie superiori	Altro	Totale
Ingegneria civile e Architettura	0,8	3,2	27,8	1,6	38,1	3,2	1,6	0,0	0,8	4,0	0,0	0,0	3,2	4,8	11,1	100,0
Ingegneria dell'informazione	0,0	5,2	27,8	7,0	38,3	5,2	3,5	0,0	0,0	6,1	0,0	0,0	0,9	0,9	5,2	100,0
Ingegneria industriale	1,8	3,0	26,2	6,7	35,4	5,5	7,9	0,0	0,6	6,1	0,6	0,0	0,0	2,4	3,7	100,0
Scienze agrarie	0,7	1,5	11,9	5,2	51,5	6,7	1,5	0,0	0,0	10,4	0,7	0,0	1,5	1,5	6,7	100,0
Scienze biologiche	3,9	1,6	12,4	5,9	48,1	5,2	1,3	0,0	0,3	8,0	0,0	0,5	2,1	1,3	9,6	100,0
Scienze chimiche	4,1	2,1	22,3	10,9	36,8	5,7	2,6	0,0	0,0	4,7	0,0	0,0	0,5	1,0	9,3	100,0
Scienze della terra	0,0	5,1	20,4	3,6	40,1	3,6	2,2	0,0	0,0	9,5	0,7	0,7	3,6	3,6	6,6	100,0
Scienze dell'antichità e filologico-letterarie	0,5	2,5	13,6	2,0	26,8	4,5	3,0	1,0	0,0	1,0	0,0	0,0	7,6	19,7	17,7	100,0
Scienze delle attività motorie e delle discipline sportive	12,5	0,0	0,0	0,0	62,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	6,3	0,0	18,8	100,0
Scienze economiche (aziendali)	2,1	10,6	25,5	2,1	25,5	6,4	17,0	4,3	0,0	0,0	0,0	0,0	0,0	2,1	4,3	100,0
Scienze economiche (socio-politiche)	2,0	6,1	19,2	5,1	39,4	7,1	4,0	2,0	0,0	4,0	0,0	0,0	1,0	0,0	10,1	100,0
Scienze fisiche	0,5	1,4	13,5	2,8	50,2	11,2	2,3	0,0	0,0	10,7	0,0	0,0	0,5	3,7	3,3	100,0
Scienze giuridiche	4,4	15,9	17,7	1,8	30,1	2,7	8,0	0,9	0,9	2,7	0,9	0,0	1,8	0,9	11,5	100,0
Scienze informatiche	0,0	11,1	22,2	11,1	16,7	11,1	5,6	0,0	0,0	22,2	0,0	0,0	0,0	0,0	0,0	100,0
Scienze matematiche	1,8	0,0	7,0	3,5	61,4	7,0	3,5	0,0	0,0	3,5	0,0	0,0	3,5	5,3	3,5	100,0
Scienze mediche	27,5	1,8	5,3	3,8	36,8	3,8	3,5	0,0	0,0	6,1	0,3	0,0	1,2	2,0	7,9	100,0
Scienze pedagogiche	3,4	0,0	0,0	0,0	31,0	3,4	3,4	0,0	0,0	0,0	0,0	0,0	0,0	20,7	37,9	100,0
Scienze politiche e sociali	2,0	5,1	18,2	6,1	40,4	6,1	4,0	0,0	0,0	2,0	0,0	1,0	0,0	1,0	14,1	100,0
Scienze psicologiche, geografiche e demotnoantropologiche	1,4	0,0	10,8	4,1	47,3	6,8	6,8	0,0	0,0	1,4	0,0	0,0	0,0	5,4	16,2	100,0
Scienze statistiche	0,0	10,5	21,1	0,0	47,4	5,3	5,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	10,5	100,0
Scienze storiche e filosofiche	0,8	4,7	20,5	1,6	30,7	3,1	0,0	0,8	0,0	1,6	0,0	0,0	6,3	15,0	15,0	100,0
Scienze storico-artistiche	0,0	7,8	17,6	2,0	29,4	3,9	2,0	0,0	0,0	0,0	0,0	0,0	5,9	11,8	19,6	100,0
Scienze veterinarie	7,9	2,6	5,3	5,3	50,0	7,9	7,9	0,0	0,0	2,6	0,0	0,0	0,0	0,0	10,5	100,0
<b>Totale</b>	<b>5,1</b>	<b>3,5</b>	<b>15,9</b>	<b>4,5</b>	<b>39,9</b>	<b>5,4</b>	<b>3,4</b>	<b>0,3</b>	<b>0,1</b>	<b>5,5</b>	<b>0,2</b>	<b>0,1</b>	<b>2,1</b>	<b>4,3</b>	<b>9,6</b>	<b>100,0</b>

**Tabella C.A019a**

Le attività di ricerca svolte nella posizione attuale sono coerenti con la formazione acquisita nel Corso di Dottorato in cui ha conseguito il titolo? ((A14>=4 e A14<=12) o A16<=2)

Valori assoluti

Area disciplinare	Pienamente coerente	Parzialmente coerente	Non coerente	Totale
Ingegneria civile e Architettura	34	23	2	59
Ingegneria dell'informazione	42	21	2	65
Ingegneria industriale	70	26	3	99
Scienze agrarie	72	25	3	100
Scienze biologiche	168	69	23	260
Scienze chimiche	61	45	8	114
Scienze della terra	52	26	5	83
Scienze dell'antichità e filologico-letterarie	57	12	1	70
Scienze delle attività motorie e delle discipline sportive	9	1	0	10
Scienze economiche (aziendali)	16	10	0	26
Scienze economiche (socio-politiche)	35	23	2	60
Scienze fisiche	111	40	8	159
Scienze giuridiche	39	14	0	53
Scienze informatiche	12	0	0	12
Scienze matematiche	28	14	2	44
Scienze mediche	113	59	8	180
Scienze pedagogiche	8	3	0	11
Scienze politiche e sociali	46	13	1	60
Scienze psicologiche, geografiche e demograficoantropologiche	32	12	2	46
Scienze statistiche	5	4	0	9
Scienze storiche e filosofiche	37	9	2	48
Scienze storico-artistiche	16	1	0	17
Scienze veterinarie	21	5	1	27
<b>Totale</b>	<b>1084</b>	<b>455</b>	<b>73</b>	<b>1612</b>

**Tabella C.A019b**

Le attività di ricerca svolte nella posizione attuale sono coerenti con la formazione acquisita nel Corso di Dottorato in cui ha conseguito il titolo? ((A14>=4 e A14<=12) o A16<=2)

Valori percentuali di riga

Area disciplinare	Pienamente coerente	Parzialmente coerente	Non coerente	Totale
Ingegneria civile e Architettura	57,6	39,0	3,4	100,0
Ingegneria dell'informazione	64,6	32,3	3,1	100,0
Ingegneria industriale	70,7	26,3	3,0	100,0
Scienze agrarie	72,0	25,0	3,0	100,0
Scienze biologiche	64,6	26,5	8,8	100,0
Scienze chimiche	53,5	39,5	7,0	100,0
Scienze della terra	62,7	31,3	6,0	100,0
Scienze dell'antichità e filologico-letterarie	81,4	17,1	1,4	100,0
Scienze delle attività motorie e delle discipline sportive	90,0	10,0	0,0	100,0
Scienze economiche (aziendali)	61,5	38,5	0,0	100,0
Scienze economiche (socio-politiche)	58,3	38,3	3,3	100,0
Scienze fisiche	69,8	25,2	5,0	100,0
Scienze giuridiche	73,6	26,4	0,0	100,0
Scienze informatiche	100,0	0,0	0,0	100,0
Scienze matematiche	63,6	31,8	4,5	100,0
Scienze mediche	62,8	32,8	4,4	100,0
Scienze pedagogiche	72,7	27,3	0,0	100,0
Scienze politiche e sociali	76,7	21,7	1,7	100,0
Scienze psicologiche, geografiche e demograficoantropologiche	69,6	26,1	4,3	100,0
Scienze statistiche	55,6	44,4	0,0	100,0
Scienze storiche e filosofiche	77,1	18,8	4,2	100,0
Scienze storico-artistiche	94,1	5,9	0,0	100,0
Scienze veterinarie	77,8	18,5	3,7	100,0
<b>Totale</b>	<b>67,2</b>	<b>28,2</b>	<b>4,5</b>	<b>100,0</b>

Tabella C.A029a

Quanto le è utile sul lavoro la formazione ricevuta durante il corso di dottorato?

Valori assoluti

Area disciplinare	È fondamentale per le mansioni che svolgo	È utile per il mio approccio al lavoro, anche se non ne faccio	Ne faccio un uso piuttosto limitato	Non ne faccio alcun uso	Totale
Ingegneria civile e Architettura	56	76	20	21	173
Ingegneria dell'informazione	51	46	16	9	122
Ingegneria industriale	86	60	16	11	173
Scienze agrarie	71	43	14	13	141
Scienze biologiche	167	148	45	29	389
Scienze chimiche	70	81	23	16	190
Scienze della terra	60	52	23	10	145
Scienze dell'antichità e filologico-letterarie	72	87	27	24	210
Scienze delle attività motorie e delle discipline sportive	8	7	1	2	18
Scienze economiche (aziendali)	21	24	15	1	61
Scienze economiche (socio-politiche)	47	35	16	13	111
Scienze fisiche	104	70	19	17	210
Scienze giuridiche	65	79	21	11	176
Scienze informatiche	10	4	0	3	17
Scienze matematiche	32	16	3	4	55
Scienze mediche	103	167	74	40	384
Scienze pedagogiche	14	9	6	3	32
Scienze politiche e sociali	45	45	7	13	110
Scienze psicologiche, geografiche e demoeoantropologiche	35	30	16	6	87
Scienze statistiche	10	7	2	1	20
Scienze storiche e filosofiche	43	52	12	22	129
Scienze storico-artistiche	16	19	7	12	54
Scienze veterinarie	20	14	5	2	41
<b>Totale</b>	<b>1206</b>	<b>1171</b>	<b>388</b>	<b>283</b>	<b>3048</b>

Tabella C.A029b

Quanto le è utile sul lavoro la formazione ricevuta durante il corso di dottorato?

Valori percentuali di riga

Area disciplinare	È fondamentale per le mansioni che svolgo	È utile per il mio approccio al lavoro, anche se non ne faccio	Ne faccio un uso piuttosto limitato	Non ne faccio alcun uso	Totale
Ingegneria civile e Architettura	32,4	43,9	11,6	12,1	100,0
Ingegneria dell'informazione	41,8	37,7	13,1	7,4	100,0
Ingegneria industriale	49,7	34,7	9,2	6,4	100,0
Scienze agrarie	50,4	30,5	9,9	9,2	100,0
Scienze biologiche	42,9	38,0	11,6	7,5	100,0
Scienze chimiche	36,8	42,6	12,1	8,4	100,0
Scienze della terra	41,4	35,9	15,9	6,9	100,0
Scienze dell'antichità e filologico-letterarie	34,3	41,4	12,9	11,4	100,0
Scienze delle attività motorie e delle discipline sportive	44,4	38,9	5,6	11,1	100,0
Scienze economiche (aziendali)	34,4	39,3	24,6	1,6	100,0
Scienze economiche (socio-politiche)	42,3	31,5	14,4	11,7	100,0
Scienze fisiche	49,5	33,3	9,0	8,1	100,0
Scienze giuridiche	36,9	44,9	11,9	6,3	100,0
Scienze informatiche	58,8	23,5		17,6	100,0
Scienze matematiche	58,2	29,1	5,5	7,3	100,0
Scienze mediche	26,8	43,5	19,3	10,4	100,0
Scienze pedagogiche	43,8	28,1	18,8	9,4	100,0
Scienze politiche e sociali	40,9	40,9	6,4	11,8	100,0
Scienze psicologiche, geografiche e demoeoantropologiche	40,2	34,5	18,4	6,9	100,0
Scienze statistiche	50,0	35,0	10,0	5,0	100,0
Scienze storiche e filosofiche	33,3	40,3	9,3	17,1	100,0
Scienze storico-artistiche	29,6	35,2	13,0	22,2	100,0
Scienze veterinarie	48,8	34,1	12,2	4,9	100,0
<b>Totale</b>	<b>39,6</b>	<b>38,4</b>	<b>12,7</b>	<b>9,3</b>	<b>100,0</b>

Tabella C.A033a

Negli ultimi tre mesi Lei ha cercato attivamente un nuovo lavoro?

Valori assoluti

Area disciplinare	Sì	No	Totale
Ingegneria civile e Architettura	53	114	167
Ingegneria dell'informazione	37	88	125
Ingegneria industriale	38	132	170
Scienze agrarie	39	98	137
Scienze biologiche	125	265	390
Scienze chimiche	63	135	198
Scienze della terra	57	89	146
Scienze dell'antichità e filologico-letterarie	57	156	213
Scienze delle attività motorie e delle discipline sportive	4	13	17
Scienze economiche (aziendali)	12	47	59
Scienze economiche (socio-politiche)	42	68	110
Scienze fisiche	61	152	213
Scienze giuridiche	28	153	181
Scienze informatiche	7	10	17
Scienze matematiche	15	41	56
Scienze mediche	94	293	387
Scienze pedagogiche	4	28	32
Scienze politiche e sociali	43	71	114
Scienze psicologiche, geografiche e demograficoantropologiche	30	57	87
Scienze statistiche	1	18	19
Scienze storiche e filosofiche	39	91	130
Scienze storico-artistiche	20	36	56
Scienze veterinarie	14	28	42
<b>Totale</b>	<b>883</b>	<b>2183</b>	<b>3066</b>

Tabella C.A033b

Negli ultimi tre mesi Lei ha cercato attivamente un nuovo lavoro?

Valori percentuali di riga

Area disciplinare	Sì	No	Totale
Ingegneria civile e Architettura	31,7	68,3	100,0
Ingegneria dell'informazione	29,6	70,4	100,0
Ingegneria industriale	22,4	77,6	100,0
Scienze agrarie	28,5	71,5	100,0
Scienze biologiche	32,1	67,9	100,0
Scienze chimiche	31,8	68,2	100,0
Scienze della terra	39,0	61,0	100,0
Scienze dell'antichità e filologico-letterarie	26,8	73,2	100,0
Scienze delle attività motorie e delle discipline sportive	23,5	76,5	100,0
Scienze economiche (aziendali)	20,3	79,7	100,0
Scienze economiche (socio-politiche)	38,2	61,8	100,0
Scienze fisiche	28,6	71,4	100,0
Scienze giuridiche	15,5	84,5	100,0
Scienze informatiche	41,2	58,8	100,0
Scienze matematiche	26,8	73,2	100,0
Scienze mediche	24,3	75,7	100,0
Scienze pedagogiche	12,5	87,5	100,0
Scienze politiche e sociali	37,7	62,3	100,0
Scienze psicologiche, geografiche e demograficoantropologiche	34,5	65,5	100,0
Scienze statistiche	5,3	94,7	100,0
Scienze storiche e filosofiche	30,0	70,0	100,0
Scienze storico-artistiche	35,7	64,3	100,0
Scienze veterinarie	33,3	66,7	100,0
<b>Totale</b>	<b>28,8</b>	<b>71,2</b>	<b>100,0</b>

## Tabelle per area disciplinare

*Tabella C.Z03c*

*Di seguito Le vengono elencati una serie di aspetti della formazione ricevuta durante il Corso di Dottorato. Per ciascuno mi dovrebbe dire quanto è soddisfatto, su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo).*

	N	Media	Mediana	Moda	SQM	Asimmetria	Curtosi	Quartile 1	Quartile 3
1. Contenuto degli insegnamenti	4207	5,88	6	8	2,49	-0,46	-0,66	4	8
2. Qualità della didattica del personale docente	4206	6,31	7	8	2,46	-0,68	-0,34	5	8
3. Qualità scientifica di professori/ricercatori	4211	7,32	8	8	2,08	-1,10	1,08	6	9
4. Sviluppo/addestramento della capacità di fare ricerca	4207	6,58	7	8	2,48	-0,66	-0,33	5	8
5. Possibilità di pubblicare (articoli, volumi, ...)	4210	6,08	7	8	2,69	-0,44	-0,80	4	8

*Tabella C.A07c*

	N	Media	Mediana	Moda	SQM	Asimmetria	Curtosi	Quartile 1	Quartile 3
Dopo quanti mesi dal conseguimento del titolo ha trovato il PRIMO lavoro iniziato dopo tale conseguimento?	2293	3,73	1	0	5,37	1,76	2,85	0	

*Tabelle C.A20c-C.A23c*

*Dopo il conseguimento del titolo di Dottore di Ricerca ((A14>=4 e A14<=12) o A16<=2)*

	N	Mean	Mediana	Mode	SQM	Asimmetria	Curtosi	Quartile 1	Quartile 3
A20A - Indichi il numero di lavori da Lei pubblicati su riviste Scientifiche a diffusione nazionale	1461	1,61	0	0	2,92	3,54	17,60	0	2
A20B - Indichi il numero delle Sue monografie pubblicate da editore nazionale	1385	0,30	0	0	0,83	4,99	38,54	0	0
A21A - Indichi il numero di lavori da Lei pubblicati su riviste Scientifiche a diffusione internazionale	1573	3,92	3	1	5,88	7,13	82,77	1	5
A21B - Indichi il numero delle Sue monografie pubblicate da editore internazionale	1325	0,40	0	0	2,52	20,59	562,31	0	0
A22 - A quanti progetti di ricerca nazionali ha partecipato	1455	1,28	1	1	1,93	16,53	451,18	0	2
A23 - A quanti progetti di ricerca internazionali (cioè finanziati da enti pubblici europei o comunque non italiani) ha partecipato	1409	0,80	0	0	1,26	4,01	29,47	0	1

Tabella C.A30c

Può esprimere su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo) il suo livello di soddisfazione per i seguenti aspetti del lavoro attuale?

	N	Media	Mediana	Mode	SQM	Asimmetria	Curtosi	Quartile 1	Quartile 3
Livello di retribuzione attuale	3351	5,60	6	7	2,33	-0,39	-0,59	4	7
Prospettive di guadagno	3347	5,04	5	6	2,55	-0,07	-0,98	3	7
Possibilità di carriera	3343	4,84	5	1	2,62	0,01	-1,07	3	7
Stabilità/sicurezza sul lavoro	3347	4,72	5	1	3,05	0,22	-1,29	2	7
Coerenza con studi fatti	3349	7,44	8	8	2,25	-1,02	0,63	6	9
Utilizzo delle competenze acquisite	3349	7,42	8	8	2,14	-1,04	0,89	6	9
Acquisizione di professionalità	3344	7,40	8	8	1,99	-1,06	1,26	6	9
Rispondenza ai propri interessi culturali	3346	7,64	8	8	2,08	-1,18	1,35	7	9
Indipendenza e autonomia sul lavoro	3344	7,43	8	8	2,08	-1,09	1,14	6	9
Coinvolgimento nei processi decisionali	3343	6,49	7	8	2,43	-0,71	-0,20	5	8
Flessibilità dell'orario di lavoro	3345	7,33	8	8	2,39	-1,05	0,51	6	9
Tempo libero	3347	5,89	6	6	2,41	-0,28	-0,67	4	8
Localizzazione sede di lavoro	3346	7,12	8	8	2,35	-0,79	0,01	6	9
Rapporti coi colleghi	3339	7,67	8	8	1,84	-1,12	1,61	7	9
Rispetto alle aspettative iniziali	3339	6,40	7	8	2,32	-0,67	-0,12	5	8
Rispetto alle mansioni svolte	3331	6,92	7	8	2,04	-0,92	0,78	6	8
Rispetto al titolo di Dottore di Ricerca	3326	5,98	6	8	2,73	-0,44	-0,87	4	8
E nel suo complesso	3324	6,75	7	7	1,79	-0,81	0,93	6	8

Tabella C.B30c

Su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo) quanto sono importanti per Lei i seguenti aspetti nella ricerca di un lavoro?

	N	Media	Mediana	Mode	SQM	Asimmetria	Curtosi	Quartile 1	Quartile 3
Prospettive di guadagno	310	7,35	7	8	1,68	-0,73	1,38	6	8
Possibilità di carriera	310	7,72	8	8	1,74	-0,76	0,90	7	9
Stabilità/sicurezza sul lavoro	311	8,54	9	10	1,60	-1,48	3,36	8	10
Coerenza con studi fatti	309	7,83	8	8	2,02	-1,36	2,29	7	9
Utilizzo delle competenze acquisite	309	8,00	8	10	1,92	-1,41	2,71	7	10
Acquisizione di professionalità	307	8,59	9	10	1,40	-1,27	2,79	8	10
Rispondenza ai propri interessi culturali	308	8,16	8	10	1,76	-1,39	3,04	7	10
Indipendenza e autonomia sul lavoro	307	7,78	8	8	1,79	-1,21	2,43	7	9
Coinvolgimento nei processi decisionali	308	7,68	8	8	1,76	-1,00	1,74	7	9
Flessibilità dell'orario di lavoro	309	7,08	7	6	2,20	-0,76	0,64	6	9
Tempo libero	308	6,74	7	7	2,10	-0,74	0,81	6	8
Localizzazione sede di lavoro	307	6,39	7	7	2,49	-0,74	-0,18	5	8
Rapporti coi colleghi	307	7,34	8	8	2,09	-1,09	1,34	6	9
Riconoscimento del titolo di Dottore di Ricerca	310	7,52	8	10	2,43	-1,08	0,72	6	10

Tabella C.C03c

Su una scala da 1 a 10 (dove 1 = per niente e 10 = moltissimo) quanto sono importanti per Lei i seguenti aspetti nella ricerca di un lavoro?

	N	Media	Mediana	Mode	SQM	Asimmetria	Curtosi	Quartile 1	Quartile 3
Prospettive di guadagno	99	6,90	7	8	2,02	-0,57	0,25	6	8
Possibilità di carriera	99	7,42	8	8	2,15	-0,82	0,35	6	9
Stabilità/sicurezza sul lavoro	97	8,49	9	10	1,53	-0,72	-0,27	7,5	10
Coerenza con studi fatti	99	8,30	9	10	1,63	-0,55	-0,92	7	10
Utilizzo delle competenze acquisite	99	8,54	9	10	1,43	-0,71	-0,25	8	10
Acquisizione di professionalità	99	8,52	9	10	1,62	-1,52	3,60	8	10
Rispondenza ai propri interessi culturali	99	8,51	9	10	1,49	-0,66	-0,62	7	10
Indipendenza e autonomia sul lavoro	99	7,91	8	8	1,49	-0,20	-0,77	7	9
Coinvolgimento nei processi decisionali	99	7,46	8	8	1,82	-0,17	-0,99	6	9
Flessibilità dell'orario di lavoro	99	7,22	7	7	1,99	-0,71	0,68	6	9
Tempo libero	99	6,78	7	7	2,27	-0,70	0,25	5	8
Localizzazione sede di lavoro	99	6,60	7	8	2,64	-0,73	-0,42	5	8
Rapporti coi colleghi	98	7,51	8	8	2,14	-1,48	2,31	7	9
Riconoscimento del titolo di Dottore di Ricerca	98	8,19	9	10	2,07	-1,29	1,48	7	10